

DOCUMENT RESUME

ED 065 008

FL 003 425

AUTHOR Lehmann, W. P.; Stachowitz, R. A.
TITLE Development of German-English Machine Translation System.
INSTITUTION Texas Univ., Austin. Linguistics Research Center.
SPONS AGENCY Rome Air Development Center, Griffiss AFB, N.Y.
REPORT NO RADC-TR-72-74
PUB DATE Apr 72
NOTE 103p.
EDRS PRICE MF-\$0.65 HC-\$6.58
DESCRIPTORS Classification; *Computational Linguistics; *English; Form Classes (Languages); *German; Language; *Linguistic Theory; *Machine Translation; Semantics; Structural Linguistics; Syntax; Systems Development; Translation; Verbs

ABSTRACT

This report deals with advances in linguistic analysis and programing. The theoretical theses for this work were presented in the first annual report. This second report concentrates on progress in the descriptive analysis of German and English, and on computer programs which have been developed during the year. Linguistic work during the period was directed at expanding the dictionaries for both German and English. The number of lexical items coded was increased, as were the features associated with each item. The coding involved complex problems of linguistic description, such as treatment of adverbs and extended forms of verbs. Since adverbs had not been adequately classified in any existing grammars of German or English, a new classificatory schema is being developed. The extended forms of German verbs have been treated in a recent monograph, which has served as the basis for the lexical analysis of these constructions within the German-English MT System. These extended forms are particularly difficult in having a small group of verbs used primarily to depict the verbal component and nouns combined with them to carry the semantic portion of the expression.
(Author/RL)

ED 065008

RADC-TR-72-74
Technical Report
April 1972



DEVELOPMENT OF GERMAN-ENGLISH
MACHINE TRANSLATION SYSTEM
The University of Texas at Austin

Approved for public release;
distribution unlimited.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

Rome Air Development Center
Air Force Systems Command
Griffiss Air Force Base, New York

FL003 425

DEVELOPMENT OF GERMAN-ENGLISH
MACHINE TRANSLATION SYSTEM

W. P. Lehmann
R. A. Stachowitz

The University of Texas at Austin

Approved for public release;
distribution unlimited.

FOREWORD

This technical report was prepared by University of Texas at Austin, Linguistics Research Center, Austin TX, under Contract F30602-70-C-0118, Job Order 45940000. The work period covered is 1 February 1971 through 31 January 1972. Mr. Charles S. Bond, Jr. (IRDT) was the RADC project engineer.

This technical report has been reviewed by the Office of Information (OI) and is releasable to the National Technical Information Service (NTIS).

This technical report has been reviewed and is approved.

Approved:


CHARLES S. BOND Jr., Sgt., USAF
Technical Evaluator

Approved:

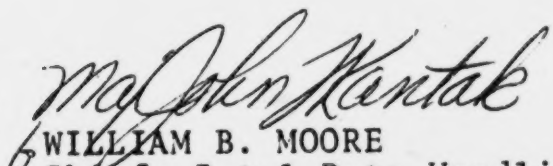

WILLIAM B. MOORE
Chief, Intel Data Handling Branch
Intel and Recon Division

TABLE OF CONTENTS

<i>ABSTRACT</i>	3
 <i>LINGUISTICS</i>	
1.1 The New English Verb List	4
1.2 New Format for German-English Verb List	10
1.3 German Verb-Noun and Verb-Prepositional Phrases	13
1.4 Feature/Subscript System for Adverbs, Prepositions, and Conjunctions	21
 <i>PROGRAMMING</i>	
2.1 Grammar Conversion	29
2.2 Systems Programs	39
2.3 Supporting Programs	47
 <i>APPENDIX A— Morphology: Trees and Rules</i>	
	64
 <i>APPENDIX B— Glossary of Subscript Definitions</i>	
	83
<i>REFERENCES</i>	97

ABSTRACT

The second annual report on work performed under contract F30602-70-C-0118 DEVELOPMENT OF GERMAN-ENGLISH MACHINE TRANSLATION SYSTEM deals with advances in linguistic analysis and programming. The theoretical theses for this work were presented in the first annual report. This second report accordingly concentrates on progress in the descriptive analysis of German and English, and on computer programs which have been developed during the year.

Linguistic work during the period was directed at expanding the dictionaries for both German and English. The number of lexical items coded was increased, as were the features noted for each item. The coding involved the linguists in some of the most complex problems of linguistic description, such as the treatment of adverbs and extended forms of verbs. As adverbs have not been adequately classified in any of the grammars, a new system of classification has had to be developed. Fortunately the extended forms of German verbs have been treated in a recent monograph, which has served as a start for our own lexical analysis. These extended forms are particularly difficult in having a small group of verbs used primarily to depict the verbal component and nouns combined with them to carry the semantic portion of the expression.

Imaginative programs have enabled us to identify the specific problems encountered in our texts. Others among those reported here are representative of the massive set of programs necessary to manage the deep as well as the surface structures of language.

LINGUISTICS

1.1 The New English Verb List

(For convenience in identifying the referents of linguistic subscripts, an alphabetized list with definitions is provided in APPENDIX B, p. 83.)

This initial portion of the second annual report is a general review of the feature/subscript system as exemplified in the revision of the English Verb List. Readers already familiar with this system and its conventions may prefer to begin with Section 1.2 New Format for German-English Verb list, p. 10.

1.1.1 Previous Work

The framework for the New English Verb List (NEV) was the English Verb List (EVL) which had earlier been prepared at LRC. The information in this list was limited to the description of the verbs and their objects. That is, verbs and verb-particle combinations were identified as being transitive or intransitive, and prepositional objects as well as adverbial prepositions were listed where applicable. This information had been obtained from *The Advanced Learner's Dictionary of Current English* by Hornby, Gatenby, and Wakefield [1]. The appropriate Verb Pattern numbers from this dictionary were listed with each entry, as was the identification of permissible objects as human, animal, inanimate, and abstract.

There were three basic tasks involved in compiling the new listing. The information in EVL had to be checked and mistakes corrected; additional information had to be supplied for each corrected verb; and new, complete entries had to be made for verbs given in *Webster's Seventh New Collegiate Dictionary* [3] but not listed in the Hornby dictionary. The present verb list, considerably revised and amplified from the original EVL, contains subject, object, and adverbial information as well as more detailed descriptions of the verbs *per se*.

1.1.2 The Verb Entry

Currently, NEV indicates for each verb whether it is a one-word entry or a two-word entry. The verb *fall*, for example, would be written as FALL 1, but a variation would be FALL BACK 2. Verbs consisting of two-word entries are composed of transitive verbs with adverbial prepositions, or of intransitive verbs with adverbial particles. An example of the former is LOOK UP 2 in *He looked (up) the number (up)*. An example of the latter is

GIVE UP 2 in *We gave up*.

In instances where a verb has two acceptable spellings, the preferred or American spelling is given first, as EXTOL, SP. EXTOLL 1.

1.1.3 Verb Description

Short descriptors, largely acronymic, are used to define the nature of the verbs. VT (Verb Transitive) means that the verb takes an object, whereas VI (Verb Intransitive) indicates that the verb has no object. VR (Verb Reflexive) is used when the object of the verb is a reflexive pronoun. VR presupposes that the verb is transitive, thus VT is not necessary. VTC (Transitive with Cognate) indicates that the verb can take only semantically cognate objects, which would include *dance a jig* as well as *dance a dance*. NP (No Passive) is used only for VT and VTC verbs. NG marks verbs which have no progressive form.

At least one of the following descriptors must be used with a verb: VT, VI, VR, VTC. The additional forms, NP and NG, cannot stand alone but are given only when necessary. They would be separated from any other symbol by a comma, e.g., TY(VI) and TY(VT, NG). (TY is the constant which references this category of descriptors.)

1.1.3.1 Subject—Form

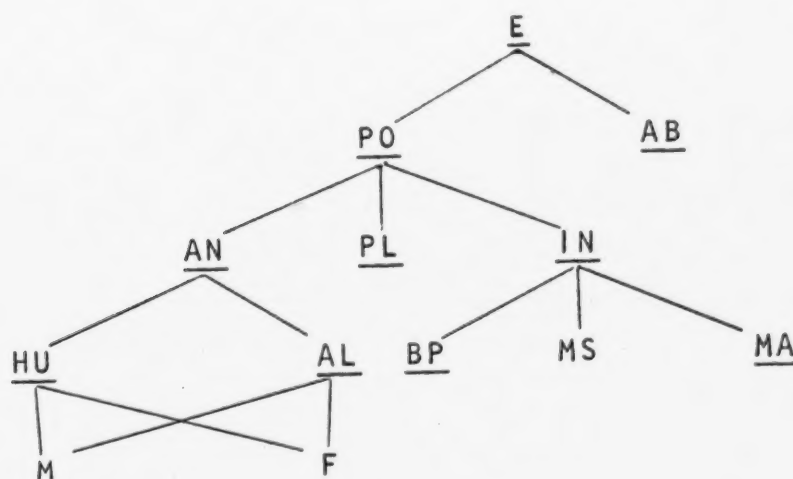
Subjects are described in two different ways, i.e., according to Form, FS, and Type, TS. The possible descriptors for form are MI, TH, GR, FT, and NP. The first four, MI (Marked Infinitive), TH (THat-clause), GR (GeRund), and FT (For-To phrase), are used only rarely. One of these occurs only when almost any possible member of that category could reasonably serve as the subject of the verb. For example, with the verb *surprise* almost any *that*-clause could function as the subject; therefore, TH would be used, FS(TH). The NP symbol refers to Noun Phrases*. These include nouns, pronouns, and clauses introduced by *what*. Usually an NP is the only type of subject possible for the verb. In such situations, the FS is omitted entirely and the subject is automatically assumed to be derived from the NP category.

One other FS descriptors is IT. This is used only for such verbs as *rain*, *snow*, etc., as in *It rains*.

* Identical descriptor symbols with different referents may be used at different points in the descriptive feature system.

1.1.3.2 Subject—Type

The descriptors for TS (Type of Subject) elucidate the nature of the subject indicated by the presence or absence of FS. They are HU (HUMAN), AL (ANIMAL), AN (ANIMATE—both human and animal), PL (PLANT), IN (INANIMATE), BP (Body Part), MA (MACHINE), MS (MasS), PO (Physical Object), AB (ABSTRACT), M (Male), F (Female), and E (*Entia*, any noun). The following tree illustrates the relationships among them.



The heading of a node automatically includes all categories below it. All underlined categories may be used alone. M and F require the presence of HU or AL. For example, such an entry as TS(HU F) means that the subject must be a human female. MS requires the presence of IN. TS(IN MS), for example, means that the subject of the verb must be an inanimate noun which, although collective, does not allow individuation of constituents.

Some verbs require a subject which must be either collective or plural, e.g., *Fish abound in the ocean*. In such instances, the subject entry would indicate this requirement in the following manner, TS(AL CO, AL P). If the verb allows virtually any category of noun as its subject as long as the noun is collective or plural, the entry would be TS(PO CO, PO P, AB P).

If the subject indicated by FS is an MI, TH, GR, or FT, the TS descriptor is automatically AB. If, however, there is no FS indicated, the subject is automatically considered to be an NP and thus must be described according to the descriptors possible for TS. If the subject can be a combination of these constructions, an entry would resemble this example:

```

TS(AB ; HU , AB)
FS(TH , MI ; NP)
  
```

The semicolon, in essence, divides two columns. All contents of the first column must correspond, as must all contents of the

second column. If the FS descriptor were IT, no TS would be needed for explanation.

1.1.3.3 Object—Form

Objects are also described according to their form (OB) and Type (TO). The possible descriptors for OB are varied: TH (THat-clause), CL (CLause without *that*), MI (Marked Infinitive), FT (For-To phrase), I (unmarked Infinitive), ICL (Interrogative CLause), IMI (Interrogative clause containing Marked Infinitive), O (noun, pronoun, or relative clause introduced by *what*), and GR (GeRund).

Sometimes adjective or noun complements are used, e.g. *I consider the man (to be) intelligent (a fool). We beat it flat. They elected him chairman.* For such constructions OB(O + complement symbol) is used. Complement symbols are BC (*to be* relational use of *to be* before noun or adjective CoMplement), NA (Noun phrase or Adjective complement without the use of *to be*), NC (Noun phrase Complement without the use of *to be*).

Prepositions may also be included as descriptors of OB. For example, direct and indirect object combinations are written as OB(O + O, TO) and OB(O + O, FOR). (The plus sign is used to show a combination of objects.) The left side of the "+" refers to the direct object, and the right side to the indirect object with its optionally deletable preposition. Sometimes the preposition cannot be deleted, e.g., OB(O + AS) in *He regarded me as a friend*. Often the preposition alone is the appropriate descriptor. This can be the result of two situations, the use of an adverbial preposition or a prepositional object. With an adverbial preposition such as *He looked up the number* or *He looked the number up*, OB(UP) is used. As was mentioned earlier concerning adverbial prepositions, the verb itself would be LOOK UP 2. The use of a prepositional object such as *rely on* in *He relied on the dictionary* requires OB(ON). The verb would be identified as RELY 1.

When the object is simply a noun phrase, OB(O) is omitted. The object will automatically be interpreted as a noun phrase.

1.1.3.4 Object—Type

The descriptors which indicate the type or nature of the object identified by the presence or absence of OB are the same as those used to identify the type of subject, and the rules for hierarchy are the same. The one exception is the use of R, signifying that the object is Reflexive, which is used when TY (VR) is present.

When the descriptor for the form of the object is TH, CL, MI, FT, I, ICL, IMI, or GR, the type of object would be written as TO(AB). If the verb is transitive or reflexive and there is no OB indicated, the object is automatically assumed to be a noun phrase and the descriptor for TO should adequately describe it. If there were two descriptors for OB joined by a plus, the descriptive information for TO must also contain a plus, and the information on each side of the "+" in TO must correspond with its counterpart in OB. The exception to this is OB(O + AC), for the AC refers to an adjective; thus only the O must be described by TO. As in *He painted the wall green*, the descriptor needs to characterize only the noun object, not the adjective complement.

1.1.3.5 Required Adverbs

Where a verb cannot be used without a certain type of adverb, the designation RA (Required Adverb) is employed. The possible descriptors are TIM (TIME), PNC (PunCTual), DUR(DURATION), PLC (PLaCe), LOC (LOCative), DIR (DIRection), ORN (ORigin), MAN (MAN-ner), MSR (MeaSure), and AC (Adjective Complement). An example is the verb *tend*. When transitive and possessing no infinitive complement, it requires an adverb of direction, RA(DIR): *He tends toward selfishness*. RA is, of course, used with only a limited number of verbs.

1.1.3.6 Optional Adverbs

When a verb possesses the aspect of motion, this characteristic is identified with the symbol OA(DOR). This means that the verb may Optionally employ an Adverb of Direction-Origin, such as a *from* or *toward* phrase, e.g., *He brought his boat from the lake*.

1.1.4 The Hornby Patterns

As in the earlier English Verb List (EVL), the Verb Pattern number from Hornby [1] is given for each verb construction in the New English Verb List (NEV). In contrast to EVL, however, inappropriate patterns have been deleted and additional workable ones have been listed. New verb entries not given in Hornby have been tested and identified according to pattern numbers. For example, the verb *slick* takes both patterns 1 and 21, and *slick up* takes both patterns 23 and 10. *Slick* is not listed as a verb by Hornby, et al., but can be assigned pattern numbers on the basis of its definitions in *Webster's Seventh New Collegiate Dictionary*.

[3]

1.1.5 Format

Many verbs require more than one entry. Some have several different types of objects possible, and convenience may necessitate separate treatment of each of these. Also, verbs which have adverbial propositions need a separate entry as the "name" of the verb varies; for example, the verb *fasten* must have entries for *fasten*, *fasten together*, *fasten up*, and *fasten down*. Certain intransitive verbs may also need several entries, e.g., *fall*, *fall back*, *fall down*, *fall out*, *fall off*. In such circumstances, one verb may include several entries.

1.1.6 Purposes of the NEV

The information in the NEV list is to be rearranged into a subscript format which will make its contents more readily amenable to computer processing. It is hoped that with the information made available the range of possible semantic meanings for a given syntactic context may be determined.

1.2 New Format for German-English Verb List

Each German verb with its descriptor sequence is identified by a unique number preceding it. Each descriptor and translation equivalent in such an entry is printed out in subscript format in a specifically numbered line as follows:

- 0 German verb stem (including prefixes). No. of morphemes
- no. of prefixes
- 1 syntactic type of verb
- 2 syntactic form of subject (NP, interrogative clause, etc.)
- 3 semantic type of subject
- 4 underlying subject (if not identical to apparent subject)
- 5 case government (cases or prepositions)
- 6 semantic type of object
- 7 type of adverbial complement required (if any)
- 8 optional adverb of direction or origin (if permissible)
- 9 auxiliary used in perfect tenses
- 10 original GVL entry in its old format
- 20 first English translation equivalent, in new format. No.
of morphemes - no. of adpreps
- 21 syntactic type of verb (for English verb)
- 25 case or preposition government (for English verb)
- 26 semantic type of object (of English verb)
- 29 original GVL entry in its old format
- 30 German verb stem, including prefixes. No. of morphemes
- no. of prefixes
- 31 syntactic type of verb (for German verb)
- 33 semantic type of subject (for German verb)
- 35 case or preposition government (for German verb)
- 36 semantic type of object (of German verb)
- 37 label of terminological sphere of applicability [domain]
(of English verb)
- 38 selectors indicating restrictions on possible subjects
or objects (of English verb)
- 39 first English translation equivalent, in old format (with
labels and selectors, if any)
- ...
- ...
- ...

59 second English translation equivalent

79 third English translation equivalent

The preceding format exhibits the following regularities: for each German verb + descriptor sequence, lines 0-10 specify all information relevant to the German verb; lines 11-19 have been left available for subsequent addition of information. Lines 20-39 form an information-block specifying all information relevant to the first English translation equivalent. These twenty lines comprise two ten-line sub-blocks, the first containing information relating directly to the properties of the English verb, and the second containing information on the German verb it translates and any restrictions on their correspondence. Subsequent blocks of the same twenty-line size, each beginning with a line whose number is a multiple of 20, specify information relevant to other English translation equivalents. Thus, for example, every $(n \times 20)$ th line contains a different English translation equivalent, displayed in the new format. Each $(n \times 20) + 9$ th line contains the original GVL entry in its old format. Each $(n \times 20) + 19$ th line contains the old-format English translation equivalent corresponding to the new-format one in the preceding $(n \times 20)$ th line.

This arrangement of the information in blocks, with corresponding specifications in corresponding line-multiples, serves two purposes. First, it facilitates the transfer of information about the properties of the German verb from the first ten lines of the entry to the relevant portion of the translation-equivalent blocks. Second, it facilitates retrieval of subsets of the list, each containing only the minimum quantity of information needed, for the greater working efficiency of the linguists.

An example of such a bilingual verb entry follows.

Earlier version: KAPITULIER VI

KAPITULIER VI VOR JDM ETW DAT

New version: GBD9201 0 KAPITULIER

1 TY (VI)

9 AUX (H)

10 GRV 3701010KAPITULIER VI

39 CAPITULATE

59 SURRENDER

GBD9202 0 KAPITULIER
 1 TY (VI)
 5 OB (VOR2)
 6 TO (HU , *AN)
 9 AUX (H)
 10 GRV 3701020KAPITULIER VI VOR JDM ETW DAT
 39 BOW TO
 59 CAPITULATE TO

In these two entries, the first indicates that the verb *KAPITULIEREN* may be used intransitively, using the auxiliary *haben* in its perfect tense forms, and that it is to be translated as *capitulate* or *surrender*. (No information was given explicitly in Wildhagen [4] concerning the semantic type of subject required by the verb.)

The second entry differs from the first in that it allows for the usage of the verb *KAPITULIEREN* with a prepositional object which must consist of the preposition *vor* and a dative noun phrase (VOR2). The semantic type of object required is specified as either human or inanimate (HU, *AN); the translations specified are *bow to* and *capitulate to*.

1.3 German Verb-Noun and Verb-Prepositional Phrases

1.3.1 Introduction

Preparatory to encoding German verb-noun phrases and verb-prepositional phrases, a study was made of Veronika Schmidt's *Die Streckformen des deutschen Verbums* [2]. It is her contention that in the past two decades the preference for phrases over simple verbal forms has become more pronounced. The need for encoding these phrases as units is evidenced by the fact that the encoding of either the verbs or the nouns alone would not encompass all the variety of shadings and differentiations which the expanded phrases exhibit. As the following examples will show, the meaning of the verb phrase is not equal to the sum of its individual parts. By pooling several components, a composite emerges with a unique definition. The individual words have relinquished their semantic identity and function as a unit.

<u>eine Frage anschneiden</u>	=	<u>to raise a question</u>
eine Frage	=	a question
anschneiden	=	to begin to cut, to begin to carve; to cut (a loaf); to give the first cut to; [fig] to broach, to embark upon, to enter upon; to raise, etc.
<u>Abstand nehmen (von)</u>	=	<u>to refrain (from), to desist (from)</u>
Abstand	=	distance, interval; difference; contrast; discrepancy
nehmen	=	to take; seize, appropriate, capture, lay hold of; receive, accept
<u>Betrachtungen anstellen</u>	=	<u>to reflect (on, upon)</u>
Betrachtung (sing.)	=	looking on; view; contemplation; study; consideration; reflection; review, discussion
anstellen	=	to place; to stand (against); [fig] to appoint; to take on, to engage; to sign on; to employ; to turn on; to switch on; to arrange, to manage, to contrive, to bring about; to commit; to cause, to do; to make

1.3.2 The Nature of Expandable Phrases

Extended verb phrases are constructed by combining an abstract noun (the semantic vehicle of the predicate) with a verb which is mainly structural in function, and possibly a preposition with attributes. This combination functions as a unit. The verb, which may be interchangeable with another verb, represents the syntactic core, e.g.,

Idm. Schande antun

Idm. Schande bringen

Idem. Schande machen

To disgrace s.o., to bring
disgrace on s.o.

Since a special meaning has been acquired in the formation of such verb phrases, a literal translation of the noun and simple verb no longer fits. This fact is particularly prominent where expanded verbal expressions convey causative meaning, degree of intensity, or aspectual action.

Causative

- a) einen Aufschub erwirken = to succeed in getting a delay (or postponement)
- b) zur Verbesserung fuehren = to lead to improvement

Degree of Intensity

- a) von Furcht ergriffen sein = to be gripped by fear
rather than
sich fuerchten = to be afraid
- b) von Schrecken erfasst sein = to be seized with panic
rather than
erschrecken = to be frightened

Ingressive

- a) die Moeglichkeit erschliessen = to offer the opportunity
- b) zur Tat schreiten - to take action
- c) Einwaende erheben = to raise objections

Durative

- a) sich in Uebereinstimmung befinden = to be in agreement
- b) sein Gewerbe betreiben = to ply one's trade
- c) den Kampf fortsetzen = to continue the struggle

Egressive

- a) Druck ausüben = to exert pressure
- b) die Ernte einbringen = to harvest
- c) eine Verbesserung einführen = to bring about an improvement

Inchoative

- a) zu Staub werden = to turn to dust
- b) in Gang kommen = to get going, to start working
- c) in Schwierigkeiten geraten = to get into difficulties

Continuative

- a) sich in den Grenzen halten = to keep within bounds
- b) in der Schwebe bleiben = to be kept in suspense

Resultative

- a) ein Ziel erreichen = to reach a goal
- b) einen Gewinn erzielen = to make a profit
- c) seinen Willen bekommen = to have one's own way
- d) in Ordnung bringen = to put in order

Causative when related to Inchoative

- a) in die Wirklichkeit umsetzen = to realize, to materialize
- b) unter Druck setzen = to (put under) pressure
- c) mit Freude erfüllen = to fill with joy

1.3.3 The Behavior of Expandable

Verb-Noun Phrases

Encoding of German verb-noun phrases and verb-prepositional phrases was based, with modifications and corrections, on the Wildhagen *German-English Dictionary* [4]. Proverbial, colloquial, obsolete, rare and slang expressions were ignored. A set of descriptors with their appropriate values was developed and subsequently refined to accommodate all grammatical peculiarities. The following examples illustrate the major problem areas.

1.3.3.1 Possessive and/or Reflexive Variables

Because the possessive/reflexive pronouns *meine*, *deine*, *seine*, *ihre*, etc., are variable, they are represented by the symbol POR as an element of the noun phrase:

a) *seine Bequemlichkeit lieben* = to be fond of comfort:

LIEB POR BEQUEMLICHKEIT TY(VI) TS(HU) = BE FOND OF
COMFORT *

Where *jemandes* (which stands for *meine, deine, seine, ihre*, etc.) is a possessive but not a reflexive variable pronoun, its underlying function is subjective, and it is encoded as POSS.

b) *in jds. Besitz uebergehen* = to become s.o.'s possession:

UEBER GEH IN POSS BESITZ TY(VI) TS(AL, PL, IN) = BECOME
POSS POSSESSION *

POR and POSS, when governing the accusative case, are left unmarked (see a) and b) above). Likewise, where a preposition within a noun phrase governs more than one case, the accusative will not be marked; one-case prepositions are not marked for case government. In all other instances, case government is marked as follows: POR/D, POR/G, POR/N; POSS/D, POSS/G and POSS/N.

Examples:

auf seiner Meinung bestehen = to stick to one's opinion:

BE- STEH AUF POR/D MEINUNG TY(VI) TS(HU) = STICK TO POR
OPINION

jdn. in seinem Beschluss bestaerken = to encourage s.o. in his decision:

BE- STAERK IN POSS/D BESCHLUSS TY(VT) OB(A) TO(HU)
TS(HU, AB) - ENCOURAGE IN POSS DECISION

eine Sache ihres Glanzes berauben = to take the bloom off s.th.:

BE- RAUB POSS/G GLANZES TY(VT) OB(A) TO(AB) TS(HU) =
TAKE THE BLOOM OFF *

sein eigener Herr sein = to be one's own master:

SEI POR/N EIGENER HERR TY(VI) TS(HU) = BE POR OWN MASTER

jds. Ebenbild sein = to be the image of s.o.:

SEI POSS/N EBENBILD TY(VI) TS(HU) = BE THE IMAGE TY(VT
NP) OB(OF) TO(HU)

1.3.3.2 Deep-Subject Phrases

Where two grammatical constructions in the English translation equivalent are possible, two entries have been made. For example,

jdm. zum Bewusstsein kommen, dass =

* If all of the descriptors are identical for both languages, they are not repeated for the English translation.

- a) *to dawn (up)on s.o. that and*
- b) *to realize that:*

- a) KOMM ZUM BEWUSSTSEIN TY(VT NP) OB(D + TH) TO (HU + AB) FS(IT) = DAWN OB(ON, UPON + TH)

where the anticipatory *it* takes the FS (form of subject) slot in both languages; and

- b) KOMM ZUM BEWUSSTSEIN TY(VT NP) OB(D + TH) TO(HU + AB) FS(IT) DS(D) = REALIZE OB(TH) TO(AB) TS(HU)

where the type of subject in English is human. In this entry the German descriptors include the deep subject in the dative case - DS(D).

1.3.3.3 Culturally and Biologically Restricted Sex Roles

In the following examples, the descriptors Type of Object (TO) and Type of Subject (TS) have to reflect disambiguated values in order to clearly indicate restricted sex roles. (Although these examples are not likely to occur in technical materials, they are included here to illustrate semantic categories that must be taken into consideration.)

- a) *sich als Mann bewahren* = *to prove oneself a man, to prove a true man*

BE- WAEHR ALS MANN TY(VR) OB(A) TO (R) TS(HU M) =
PROVE A MAN*, PROVE A TRUE MAN TY(VI)

- b) *zur Jungfrau heranwachsen* = *to grow to womanhood*

HERAN WACHS ZUR JUNGFAU TY(VI) TS(HU F) = GROW TO
WOMANHOOD*

- c) *jdn. unter die Haube bringen* = *to marry s.o. off (a girl)*

BRING UNTER DIE HAUBE TY(VT) OB(A) TO(HU F) TS(HU) =
MARRY OFF*

- d) *jdn. zur Mutter machen* = *to get s.o. with child*

MACH ZUR MUTTER TY(VT) OB(A) TO(HU F) TS(HU M) =
GET WITH CHILD*

1.3.3.4 Reflexive Phrases

Examples a) through e) illustrate inherently intransitive verb-noun phrases which have been given Object and Type of Object slots in the German entries in order to facilitate mechanical translation. As is shown here, German reflexive objects may be in the accusative, dative, or even genitive case.

a) *sich ins Faeustchen lachen* = to laugh up one's sleeve

LACH INS FAEUSTCHEN TY(VR) OB(D) TO(R) TS(HU) = [FIG]
LAUGH UP POR SLEEVE TY(VI)

b) *sich zum Guten wenden* = to take a turn for the better

WEND ZUM GUTEN TY(VR) OB(A) TO(R) TS(AB) = TAKE A
TURN FOR THE BETTER TY(VI)

c) *sich in Kohle verwandeln* = to turn into coal, to become
carbonized

VER- WANDEL IN KOHLE TY(VR) OB(A) TO(R) TS(IN MS) =
TURN INTO COAL TY(VI), BECOME CARBONIZED

d) *sich mit Laub bedecken* = to put out leaves

BE- DECK MIT LAUB TY(VR) OB(A) TO(R) TS(PL) = PUT
OUT LEAVES TY(VI)

e) *sich von den Plaetzen erheben* = to rise from their seats

ER- HEB VON DEN PLAETZEN TY(VR) OB(A) TO(R) TS(P HU)
= RISE FROM POR SEATS TY(VI)

f) *sich eine Auffassung (ueber ...) bilden* = to form a
view (of)

BILD EINE AUFFASSUNG TY(VT,VR) OB(D;D + UEBER1) TO(R;
R + E) TS(HU) = FORM A VIEW TY(VI,VT) OB(OF)

where the transitive version requires a prepositional object.

g) *mit dem Gelde um sich werfen* = to throw one's money
around

WERF MIT DEM GELDE TY(VR) OB(UM A) TO(AB R) TS(HU) =
THROW POR MONEY AROUND TY(VI)

where *um* is an integral part of the reflexive expression and is
not separated by a "plus" symbol.

h) *es sich zur Pflicht machen zu tun* = to make it one's
duty to do

MACH ZUR PFLICHT TY(VT,VR) OB(A + D + MI) TO(IT + R +
AB) TS(HU) = MAKE POR DUTY TY(VT) OB(O + MI) TO(IT +
AB)

where an *it*, the reflexive, and a marked infinitive appear in the
German lexical entry under OB and TO.

i)

1) *sich zum Unsegen auswirken* = to lead to misfortune

AUS WIRK ZUM UNSEGEN TY(VR) OB(A) TO(R) TS(AB)
FS(NP,MI,TH) = LEAD TO MISFORTUNE TY(VI)

- 2) *sich zum Unsegen auswirken, dass* = *to lead to misfortune that*

AUS WIRK ZUM UNSEGEN TY(VT, VR) OB(A + TH) TO(R + AB)
FS(IT) = LEAD TO MISFORTUNE TY(VT)

where the first entry is an intransitive phrase with an abstract subject in the form of a noun phrase, a marked infinitive, or a *that*-clause. The second entry is a transitive version with a *that*-clause in the object slot and form of subject with an "it" value.

1.3.3.5 Pseudo-Reflexive and Non-Reflexive Phrasing

Certain phrases, which at first appear to be straightforward transitive phrases, require two entries because of different translation equivalents. Note that example a) is pseudo-reflexive in the German reading and takes an intransitive translation:

sich (od jdn.) ums Leben bringen

- a) *sich ums Leben bringen* = *to commit suicide*

BRING UMS LEBEN TY(VR) OB(A) TO(R) TS(HU) = COMMIT
SUICIDE TY(VI)

- b) *jdn. ums Leben bringen* = *to kill s.o.*

BRING UMS LEBEN TY(VT) OB(A) TO(HU) TS(HU) = KILL *

1.3.3.5 Reciprocal Phrase with Plural Subject

Certain phrases require a plural subject when reciprocity is to be expressed. *Sich* often represents *einander*, an implied (*with*) each other in the English translation.

- a) *sich die Haende reichen* = *to shake hands*

REICH DIE HAENDE TY(VT NP) OB(D) TO(RCC) TS(P HU) =
SHAKE HANDS TY(VI)

- b) *sich ein Stelldichein geben* = *to meet*

GEB EIN STELLDICHEIN TY(VT NP) OB(D) TO(RCC) TS(P HU)
= MEET TY(VI)

- c) *sich in die Differenz teilen* = *to split the difference*

TEIL IN DIE DIFFERENZ TY(VT NP) OB(A) TO(RCC) TS(P HU)
= SPLIT THE DIFFERENCE TY(VI)

1.3.3.7 Plural Subject Only

- a) *an einem Joch ziehen* = *to bear the yoke together*

ZIEH AN EINEM JOCH TY(VI) TS(P HU) = [FIG] BEAR THE YOKE
TOGETHER *

b) *in Reihen (an ...) vorbeimarschieren* = to file past
 VORBEI MARSCHIER IN REIHEN TY(VI,VT) OB(AN2) TO(HU,IN)
 TS(P HU) = FILE PAST *

1.3.3.8 + or - MIT

A number of verb-noun phrases refer to a situation which must involve more than one participant. Either the subject must be plural, as in a 1) below, or a *with-object* must accompany the verb if the subject is singular, as in b 2).

- a)
- 1) *eine Abmachung treffen* = to make an agreement
 TREFF EINE ABMACHUNG TY(VI) TS(P HU) = MAKE AN AGREE=MENT *
 - 2) *eine Abmachung mit jdm. treffen* = to make an agreement with s.o.
 TREFF EINE ABMACHUNG TY(VT) OB(MIT) TO(HU) TS(HU)
 = MAKE AN AGREEMENT OB(WITH)

- b)
- 1) *einen Wortwechsel haben* = to have words
 HAB EINEN WORTWECHSEL TY(VI) TS(P HU) = HAVE WORDS *
 - 2) *einen Wortwechsel mit jdm. haben* = to have words with s.o.
 HAB EINEN WORTWECHSEL TY(VT NP) OB(MIT) TO(HU) TS(HU)
 = HAVE WORDS OB(WITH)

1.3.3.9 Eliding of Verbs

In some instances, the surface structure of a noun phrase indicates an elided verb. This necessitates the inclusion of an extra value under Type of Object, i.e.

Abstand nehmen von ... (Suessigkeiten, Alkohol, etc.) =
to renounce, to desist from (sweets, alcohol, etc.)

although deep structure analysis would yield:

Abstand nehmen von ((dem Essen der) Suessigkeiten) =
 or *((dem Trinken des) Alkohol(s))* =
to renounce (the eating of) sweets or
to renounce (the drinking of) alcohol.

Therefore, the proper encoding of the above phrase would list the value IN with a question mark for future retrieval purposes:

NEHM ABSTAND TY(VT) OB(VON) TO(AB,IN ?) TS(HU) = RENOUNCE,
 DESIST OB(FROM)

1.4 The Feature/Subscript System for Adverbs, Prepositions, and Conjunctions

In the course of refining and enlarging the Center's dictionaries, work was begun on the coding of lists of English adverbs, prepositions, and conjunctions. Each entry is coded with its German translation equivalents and with syntactic and semo-syntactic features. For this purpose, a general classification scheme was designed which is described in the following pages.

The semantic classification of each item is indicated under the subscript TY (for "type"); this classification scheme is largely the same for adverbs, prepositions, and conjunctions. All other features indicate selection restrictions; the classification schemes for these differ slightly for the three categories.

A glossary of definitions for all subscripts in alphabetical order may be found in Appendix B.

Adverbs

1.4.1 Subscripts for Features

One-word adverbs (including those derived from adjectives and present and past participles) are given some or all of the following subscripts. (In this list, mandatory subscripts are underlined and definitions are given for each symbol.)

- TY = semantic type of adverb
- PA = Paraphrasability (relevant only for parenthetical adverbs)
- MD = modifies (the adverb may modify verbs, sentences, or NP's)
- TS = semantic type of sentence subject required (relevant only with adverbs modifying verbs and, possibly, sentences)
- TV = semantic type of verb with which the adverb may be used (relevant only with adverbs modifying verbs)
- PØS = position (pre- or post-posed; sentence initial, medial, or final)
- RC = requires complement (adverbs, clauses or phrases)
- ØC = optional complement
- TN = tense (the adverb requires that the verb occur in a specific tense(s); this subscript is not coded if the same information is contained under TY in one of the values PR, PA or FU)

Each of the subscripts in the list above is associated with one or more values describing the characteristics of the particular item being classified or its selection restrictions.

1.4.1.1 Values for "Type" (TY) when used with Adverbs (Conjunctions and Prepositions)

P = parenthetical
 DEF = definite
 IND = indefinite
 L = location, which may be specified as

STA = static
 DI-T = direction to
 DI-F = direction from

T = time, which may be specified as

PR = present
 PA = past
 FU = future
 PR-T = prior to
 SIM = simultaneous with
 PØ-T = posterior to
 PU = punctual
 DU = duration (time span answering the question "how long?", e.g., *for eleven days*)
 FR = frequency (repetitive)
 SE = sequential (SE without INC or TRM means "sequential but not initial or final", e.g., *secondly*)
 INC = incipient
 TRM = terminating
 INST = instantaneous (point in time, e.g., *at 8 P.M.*)
 EXT = extended (time span answering the question "when?", e.g., *today*)

M = modal, which may be specified as

MAN = manner
 SM = state of mind
 EV = evaluation of subject (It is ADJ of SUBJECT to INFINITIVE: He wisely did it - It was wise of him to do it.)
 CØM = comparison
 CØM PEJ = comparison pejorative
 RES = restrictive
 MØD = mode of existence

D = degree, which may be specified as

LS = lower scale

MED = medium
 H-S = higher scale
 APP = approximation
 CØM = comparison

CA = cause
 PP = purpose
 I = instrumental
 R = result
 CD = condition
 CC = concessive
 MØ = modality
 ME = measure
 S = sociative
 A = adversativity

1.4.1.2 Values for "Parenthetical" (PA)

I = *it - that* paraphrase possible: He will probably come -
 It is probable that he will come.
 W = post-sentential which-relative clause possible: He sur-
 prisingly works slowly - He works slowly which is sur-
 prising.
 C = *so ADJ that* or *to a degree that* paraphrase possible (con-
 secutive): He works surprisingly slowly - He works so
 slowly that it is surprising.
 H = *it - how* paraphrase possible: He works surprisingly slowly
 - It is surprising how slowly he works.

1.4.1.3 Values for "Modifiers" (MD)

S = sentence
 D = declarative
 Q = question
 I = imperative
 N = negated D, Q, I, or S
 V = verb
 NP = noun phrase
 AV = adverb (including PRPH)
 NU = numbers
 A = adjective
 E = equative
 P = positive
 C = comparative
 SP = superlative

1.4.1.4 Values for "Type of Subject" (TS)

P = plural (i.e., the adverb requires a plural subject or a singular subject with a *mit*-PRPH)

1.4.1.5 Values for "Position" (POS)

A = ante (= pre-posed) relevant for modifiers of NP, A, AV,
P = post (= post-posed) or NU only
I = sentence-initial
M = sentence-medial relevant for modifiers of S and V
F = sentence-final only

1.4.1.6 Values for the Subscripts "Requires Complement" (RC) and "Optional Complements" (OC)

All prepositions, spelled out

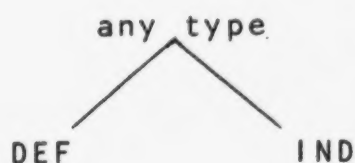
AV = any type of adverb
AV PLC = adverb of place
AV TIM = " " time
AV PNC = " " punctuality
AV DUR = " " duration
AV LOC = " " location
AV DIR = " " direction to
AV ØRN = " " origin (direction from)
AV MAN = " " manner
AV MSR = " " measure
AC = adjective complement
TH = *that*-clause
MI = marked infinitive
FT = *for-to* complement
GR = gerund
ICL = interrogative clause
IMI = interrogative adverb + marked infinitive
GG = genitive
DG = dative
AG = accusative
NPG = noun phrase

1.4.1.7 Values for "Tense" (TN)

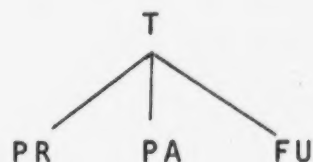
PR = occurs with verbs in present tense
PA = occurs with verbs in past tense
FU = occurs with verbs in future
PF = occurs with verbs in any perfect tense (PF may also be used together with PR, PA or FU to indicate present perfect tense, etc., as relevant)

Some values may be used in combinations, as indicated below.

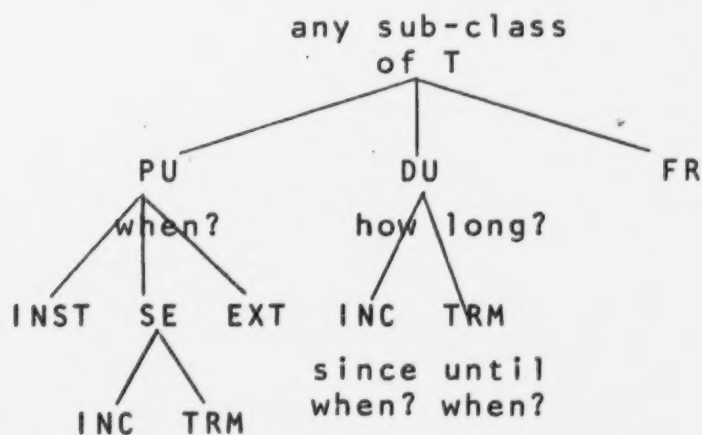
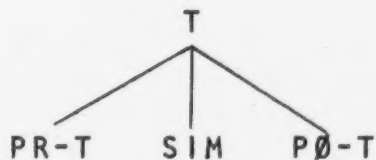
Value Trees:



where? whereto? from where?



or:



Possible Value Combinations:

Any value of TY may be combined with DEF or IND as relevant.

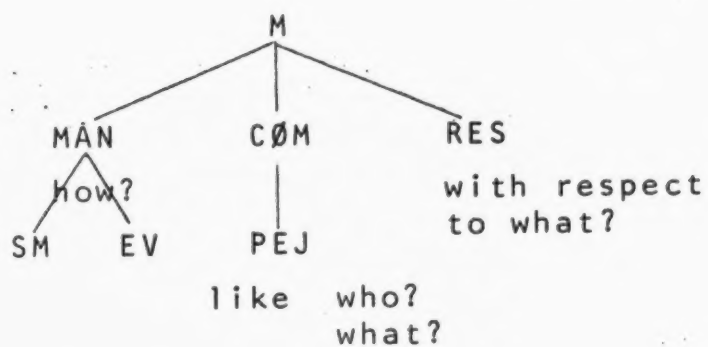
L may be combined with lower nodes (e.g. L STA).

Combine T with any of the 6 lower nodes. However, time adverbs which do not specifically indicate past, present, or future do not get the values PR, PA, or FU.

Combine any of the 6 T-dominated nodes above with the lower nodes in this tree, as relevant.

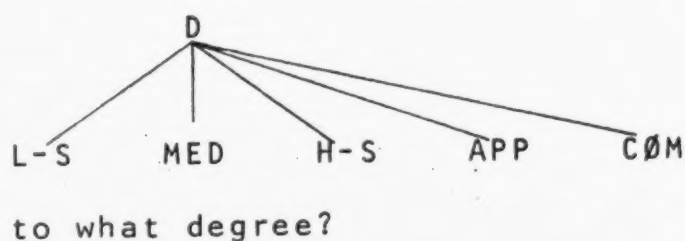
Use DU with DEF or IND to mean "limited" or "unlimited duration", respectively.

Value Trees (Cont'd):

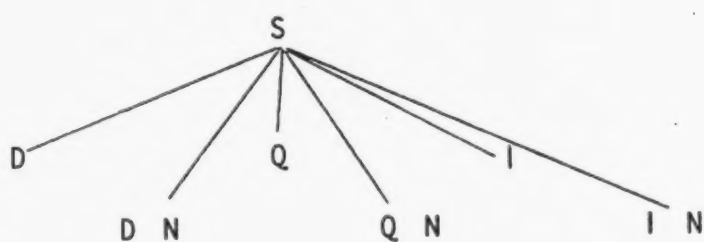


Possible Value Combinations:

Combine M with lower nodes as lines allow.



Combine D with lower nodes.



Combine S or A with any of the lower nodes as relevant.

S by itself means that the particular adverb can modify any of the 6 types of S; the same is true of A.



Conjunctions

1.4.2 Subscripts

C-1 = connects or introduces (clauses, noun phrases, verb phrases, etc.)
TY = type of conjunction

1.4.2.1 Values of "Connects or Introduces" (C-1)

MC = main clause
SC = subordinate clause
A = adjective or adverb (phrase)
N = noun (phrase)
V = verb (phrase)

1.4.2.2 Values of "Type" (TY)

CØNJ = conjunctive (takes pl. verb: *and*)
DISJ = disjunctive (takes sg. verb: *or*)

In addition, all semantic features under the subscript TY of adverbs may be used for conjunctions (cf. p. 22).

Prepositions

1.4.3 Subscripts

TY = semantic type of preposition
RC = requires complement (noun phrase or adverb)
TC = semantic type of complement
POS = position (pre- or post-posed)

1.4.3.1 Values for "Type" (TY)

Cf. TY values for adverbs, p. 22.

1.4.3.2 Values for "Requires Complement" (RC)

NP = noun phrase
AV = adverb

1.4.3.3 Values for "Semantic Type of Complement" (TC)

The possible values for this subscript are —

a) all semantic subclasses for nouns:

PO = physical object
AB = abstract
AN = animate
PL = plant
IN = inanimate
HU = human
AL = animal
NM = proper name
CO = collective (components are countable)
BP = body part
MS = mass (homogeneous)
MA = machine
QU = quantity (___ + (of) NP; e.g., *group*, as in
 a group of items)
CN = count (abstract countable nouns)
UN = unit (ADV = QUANT + ___; e.g., *mile*, as in
 five miles long)

b) all values given for the subscript TY (type) of
adverbs (cf. p.22).

1.4.3.4 Values for "Position" (POS)

PRE = pre-posed to the NP or AV
POST = post-posed to the NP or AV

In addition, a separate list is being coded of all those English items whose German translation equivalents are separable verbal prefixes with independent meaning. An example would be the English preposition *up* with the German translation *auf*, as in:

get up = *aufstehen*.

PROGRAMMING

During the second year of work performed on DEVELOPMENT OF GERMAN-ENGLISH MACHINE TRANSLATION SYSTEM, the programming effort was divided into three areas: grammar conversion programs, systems programs, and supporting programs.

2.1 Grammar Conversion

Three major items were completed and are described in the immediately following pages:

- a) The RFMS F1 English dictionaries RMD and WEBSTER were conflated into one dictionary and converted into RFMS F2 format.
- b) The RFMS F2 German dictionary was updated to include canonical forms and to associate matching rule numbers and different duplication numbers with each set of allomorphs.
- c) The RFMS F1 German transfer grammar was converted into RFMS F2 format.

2.1.1 Conflating the RMD and WEBSTER Dictionaries

A set of programs was designed to combine the two RFMS F1 loader-format dictionary grammars, RMD and WEBSTER, in order to produce one standardized RFMS F2 loader-format dictionary grammar. In combining the two dictionaries, redundant rules were eliminated.

2.1.1.1 Sequences of Operations

The following steps, described in 2.1.1.4-8 were involved in the conflation:

- a) preparation of RMD rules for conflation
- b) preparation of WEBSTER rules for conflation
- c) conflation of RMD and WEBSTER
- d) processing of WEBSTER rules not found in the intersection of RMD and WEBSTER
 - 1) assignment of the same rule numbers to allomorphs of rules
 - 2) elimination of redundant rule numbers
- e) production of rules for converting unique class names from RFMS F1 loader-format to RFMS F2 loader-format
- f) conversion of rules produced by the preceding steps from RFMS F1 loader-format to RFMS F2 loader format
- g) listing of nouns which lacked TYPE information (e.g., human [HU], abstract [AB], etc.), to facilitate the addition of such information by the linguists.

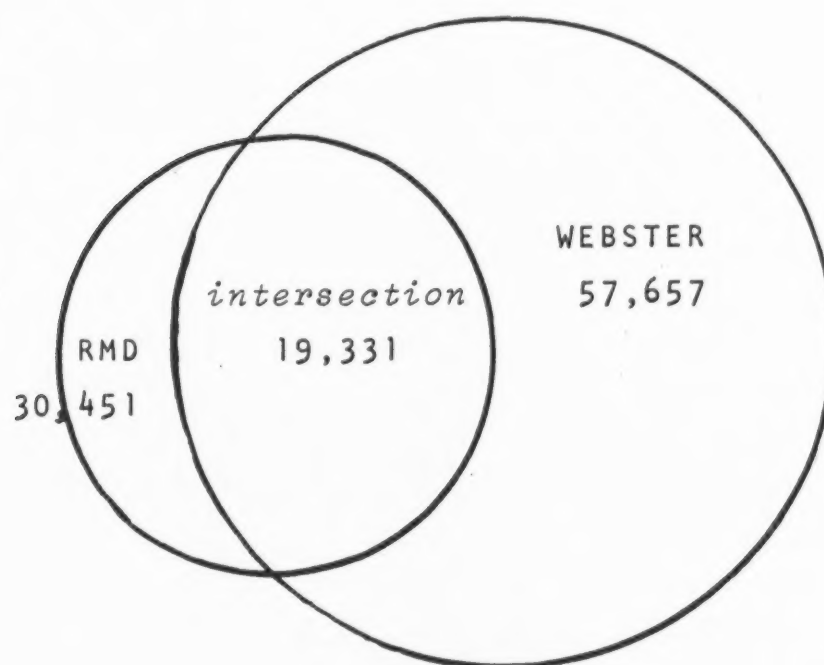
2.1.1.2 General Statistics for the Conflation

Number of rules in—

RMD before revision	42,582
WEBSTER before revision	77,412
RMD after revision	49,782
WEBSTER after revision	76,909
Input to conflation	<u>126,671</u>

Input to conflation	126,671
Intersection of RMD and WEBSTER	19,331
RMD, not in the intersection	30,451
WEBSTER, not in the intersection	57,657
NEW ENGLISH DICTIONARY	<u>107,439</u>

2.1.1.3 Diagram of Conflation



2.1.1.4 Preparation of RMD Rules for Conflation

Two programs were involved in the preparation of the RMD rules for conflation. The first program deleted the following types of rules from the input to the conflation program:

- a) rules with invalid format
- b) rules with class names specified by the linguists as "do not convert"
- c) rules with invalid class names.

Such rules were assembled according to type and printed.

Rules with multi-class names were then broken down into single-class name rules; for each additional rule—

- a) new rule numbers were assigned
- b) new rule numbers were associated with their original rule numbers and assembled for later printing
- c) information relating to each associated class name was created and attached to the class name.

In the second program involved in the preparation of RMD rules for conflation—

- a) allomorphs of rules were created, i.e., classnames and right-side terms were altered to a pseudo-WEBSTER format according to change rules created by the linguists;
- b) canonical information was attached to the class names of the rules according to specifications which the linguists supplied;
- c) rules were sorted according to the combined constituents of the right-side terms;
- d) all rules were converted to a format amenable to processing by the conflation program.

2.1.1.4.1 Statistics for RMD

Number of rules—

in the original RMD dictionary grammar	42,582
with invalid format	- 1
with class names specified by the linguists as "do not convert"	-168
before expansion of multi-class names into single-class names	42,413
after expansion of multi-class names (i.e., before generation of allomorphs)	46,530
in pseudo-WEBSTER format with canonical information attached (i.e., after generation of allomorphs)	49,782

2.1.1.5 Preparation of Rules for Conflation

Two programs were involved in the preparation of the WEBSTER rules for conflation, also. The first program deleted the following types of rules from the input to the conflation—

- a) rules with invalid format
- b) rules with the RFMS F1 loader-format item 9] BAD, which denotes an error in the rule.

These rules were assembled according to type and printed.

In accordance with information specified by the linguists, the second program attached irregular-verb canonical information to the class names of rules. All rules were converted to the format amenable to processing by the conflation program.

2.1.1.5.1 Statistics for WEBSTER

Number of rules—

in the original WEBSTER dictionary grammar	77,412
with invalid format	- 4
with RFMS F1 loader-format item 9] BAD	-499
in format amenable to conflation program	76,909

(with canonical information attached, i.e., irregular verbs)	286
---	-----

2.1.1.6 Conflation of RMD and WEBSTER

The RMD-WEBSTER Conflation program (CONFLAT) compares RMD dictionary grammar rules in RFMS F1 loader-format to WEBSTER dictionary grammar rules in RFMS F1 loader format. It produces three output files and two special purpose files based on this comparison.

CONFLAT uses two input files. The first file contains the RMD rules in RFMS F1 loader-format with a sort key consisting of:

- a) the combined constituents of the right-side terms
- b) the category symbol
- c) canonical information
- d) type information

The second file contains WEBSTER rules in essentially the same format as the RMD file. The one exception lies in the use of a

special symbol indicating the presence of canonical rather than type information in the WEBSTER rule.

The three output files from CONFLAT are as follows:

- a) a file of RMD rules with no WEBSTER equivalent. These consist of rules with sort keys stripped off.
- b) a file of WEBSTER rules with no RMD equivalent. The records on this file are copied exactly as they are on input, including the sort keys. These records are to be used for further processing.
- c) a new dictionary file (MATCH) consisting of the matched records plus RMD records without a match. (This implies that there are as many rules in this group as on the original RMD file.) When a match occurs, the type information is inserted (if it is non-zero) into the class name of the RMD record unless a special symbol occurs in the WEBSTER rule. In the latter case, the type information is inserted into the WEBSTER rule. The appropriate rule is then added to the new dictionary file.

The two special purpose files resulting from CONFLAT are used by successive programs. The first file consists of unique class names. A table is printed which lists each class name and its frequency count. The other file contains rule-number pairs which represent matches, i.e., the RMD rule number and the WEBSTER rule number where complete matches occur.

In addition to these files, CONFLAT generates a printed list which contains the RMD class and WEBSTER class associated with it whose combined right-side terms match but whose class names differ.

2.1.1.7 Processing of Rules Not Found in the Intersection of RMD and WEBSTER

A stem-stripping and comparison program (STSTRIP) was written to assign the same rule numbers to allomorphs in the file of WEBSTER rules not found in the intersection of RMD and WEBSTER (WEBNOT). For each rule in WEBNOT, STSTRIP generated a sort key containing:

- a) the combined constituents of the right-side terms with endings stripped off either according to canonical information, if available, or to rules specified by the linguists, and
- b) an indicator for linking class names in the comparison of the stripped combined right-side terms.

The STSTRIP program assigned the same rule number to each pair of rules which had matching stripped right-side terms and linked class names. Therefore, allomorphs of a rule were assigned the same rule number and redundant rule numbers were eliminated. STSTRIP printed the list of eliminated rule numbers with their corresponding replacement rule numbers.

STSTRIP discarded the sort key from the rules, producing a new file, WEBNOT1. WEBNOT1 and MATCH (the file consisting of rules which matched in the intersection of RMD and WEBSTER, plus RMD rules with no WEBSTER equivalent) were converted to RFMS F2 loader-format. This RFMS F2 version of the combined dictionary grammars is now referred to as the New English Dictionary Grammar, NEWENG.

2.1.1.7.1 Statistics for STSTRIP

Number of rules:

in WEBNOT	57,667
with no allomorphs	51,545
with allomorphs (involving 3,061 matches)	6,122
in WEBNOT1	57,667

(Note that rule numbers, not rules, were eliminated.)

2.1.1.8 Conversion Rules for Unique Class Names

A further program produced conversion rules for each unique class name found in NEWENG. Each class name consisted of codes for the following information:

- a) canonical (CAN)
- b) type (TY)
- c) category, indicating associated class information (CAT)
- d) class information concerning—
 - 1) left category symbol
 - 2) onset (ON)
 - 3) class (CL)

These codes were processed to produce conversion rules for converting NEWENG from RFMS F1 to RFMS F2 loader-format.

2.1.2 German Dictionary Grammar

The F2 German dictionary was updated to include canonical forms and to associate matching rule numbers and different duplication numbers with each set of allomorphs. This was accomplished in three steps.

Step 1 involved adding a PCA (Possible CANonical form) subscript to each rule whose right-side contained an umlaut (coded by the characters AE, OE, or UE). The values of the PCA subscript were encoded in a fixed format to resemble commands that a pseudo-string processor might execute. For example, the rule

```
C 2836  V N          * AEMT
      + GL(8)
      + GD(P)
```

was given the subscript + PCA(D.3), which has the meaning: to construct the canonical form of AEMT delete the third character from the right. This resulted in the form AMT.

Step 2 adds a CAN (CANonical form) subscript to rules specifying certain paradigmatic classes and to all irregular verb rules. For example, the rule

```
C 25598  V A          * SENSIBL
      + CV
      + IN
```

was given the subscript + CAN(1.1'E) which has the meaning: to construct the canonical form of SENSIBL, insert an E before the first character from the right. This resulted in the form SENSIBEL.

An example of an irregular verb is the rule

```
C 41701  V V          * RAENG
      + CL(15)
      + PCA(D.3)
```

This rule received the subscript + CAN(R'RING) which directs the entire word to be replaced by the word RING.

Step 3 involved the matching of rules having PCA and/or CAN subscripts with rules having neither. The match is performed first by the use of the category symbol, and second by comparing the result of executing the PCA and/or CAN subscript to the right-side of rules without these subscripts. (The CAN was used when both it and PCA occurred in the same rule.) For each such match the same rule number and a different duplication number was assigned to the rules involved. (If the match involved a rule with only a PCA, the PCA was changed to a CAN.)

The result of this step was a series of commands by which DICT UP performed the above corrections. After the update was performed, the above rules (along with the rules they matched) now look as follows:

C	2099	V N	* AMT
D	1	+ CL(5)	
		+ GD(N)	
C	2099	V N	* AEMT
D	2	+ CL(8)	
		+ GD(P)	
		+ CAN(D.3)	
C	24574	V A	* SENSIBEL
D	1	+ PV	
		+ SV(S)	
C	24574	V A	* SENSIBL
D	2	+ CV	
		+ IN	
		+ CAN(I.1'E)	
C	25909	V V	* RING
D	1	+ CL(48)	
C	25909	V V	* RAENG
D	3	+ CL(15)	
		+ PCA(D.3)	
		+ CAN(R'RING)	

During this correction process, rules originally having rule numbers 2836, 25598, and 41701 were eliminated.

2.1.3 Transfer Conversion

The conversion of the existing transfer grammars from RFMS F1 to RFMS F2 was divided into three steps. The first two steps were concerned with preparing the input. Step One extracted the rule number and left- and right-sides. Changing each left category symbol to an abbreviated form such as N, V, or A, it sorted the information by rule number. It then arranged the information into a structure having the property that any rule could be accessed directly by knowing only its rule number.

Step Two performed the same operations for the 160,000 unique interlingual substitution symbols. The information which was retained in this step was the indication of which of the names was English.

The third step was the actual program which converted each RFMS F1 transfer rule to the new RFMS F2 format.

The German transfer grammar has been successfully converted. The following are examples of the present form of entries:

C 1509	V CONFUSED	C 7107
	A CAT(N)	
	N TM(DURCHEINANDER)	
C 1515	V DEDUCE	C 37235
	A CAT(V)	
	N TM(ENTNEHM)	

(The value of the subscript CAT represents the left side of the German dictionary rule; the value of TM, the right side.)

2.2 Systems Programs

The following systems programs were constructed:

- a) Word grammar compilation
 - 1) Word grammar sort (WORD GS)
 - 2) Word grammar tree construction (WORD TC)
- b) Word Analysis (WORD A)
- c) Transfer grammar compilation
 - 1) Transfer grammar sort (TRAN GS)

These are described in the following pages.

2.2.1.1 Word Grammar Sort

Using word grammar rules in RFMS F2 loader-format, the Word Grammar Sort program (WORD GS) creates and sorts records to be used as input to WORD TC, the Word Tree Construction program. A record is created for each left- or right-side term in each rule. Thus, for a rule containing one left-side term and two right-side terms, three records would be created. A description of the types of records created follows.

Each sort record contains a number denoting its term position. Left-side terms are assigned the term number 0; right-side terms are numbered from right to left, beginning with the term number 1. The following rule shows such term numbers listed above their respective terms.

	0	3	2	1
C 25	V WORD	V LB	V NO	V RB
			B	B

(rule number) (left-side term) (right-side terms...)

For the left-side term of a word grammar rule, a sort record is created which contains a sort key consisting of:

- a) the term number 0
- b) the left category symbol
- c) all other left-side information, i.e., left-subscripts, left-operators, etc.

and a sort data area consisting of:

- a) rule number
- b) any dummy term information. (In the rule, as opposed to the record being created, this information follows the right-side terms.)

For each right-side term of a word grammar rule, a sort record is created which contains a sort key consisting of:

- a) term number (right side terms are numbered from right to left, as shown above)
- b) right category symbol
- c) all other right-side information for the term being processed, i.e., right-subscripts, right-operators, etc.

and a sort data area consisting of:

- a) rule number
- b) number of total right-side terms for the rule.

For the example shown above, the following sort records

would be created:

- a) the record for the left-side term, which contains—
 - 1) 0 (term number)
 - 2) WORD (left category symbol)
 - 3) 25 (rule number)
- b) the record for the first right-side term, containing—
 - 1) 3 (term number)
 - 2) LB (right category symbol)
 - 3) 25 (rule number)
 - 4) 3 (number of right-side terms)
- c) the record for the second right-side term, containing—
 - 1) 2 (term number)
 - 2) NO (right category symbol)
 - 3) B (right operator)
 - 4) 25 (rule number)
 - 5) 3 (number of right-side terms)
- d) the record for the third right-side term, containing—
 - 1) 1 (term number)
 - 2) RB (right category symbol)
 - 3) B (right operator)
 - 4) 25 (rule number)
 - 5) 3 (number of right-side terms).

All records created are in a format amenable to processing by the CDC 6600 SORT/MERGE routine.

WORD GS invokes the SORT/MERGE routine. Upon completion of the sort, all terms for all rules will be grouped according to term numbers.

2.2.1.2 Word Grammar Tree Construction

From the Word Grammar rule terms produced by WORD GS, the Word Tree Construction program (WORD TC) generates the internal structure of the Word Grammar which will be used in the Word Analysis program. Each entry in the input file represents a term in a Word Grammar rule.

WORD TC produces a file amenable to processing by Word Analysis. This file contains the following three tables:

- a) Subscript Packages Table containing the unique Subscript Packages (i.e., all information pertinent to the term) and their associated rule numbers;
- b) Category Symbols Table containing Category Symbols and the addresses in the Subscript Package Table where those Subscript Packages associated with successive Category Symbols are found;
- c) Category Symbols Addresses Table (located in the Category Symbols Table) containing addresses of Category Symbols for successive rule terms.

Each first term (i.e., the left-most term on the right-side of a rule) contains a pointer to its associated left-side term.

2.2.2 Word Analysis

A partial description of Word Analysis (WORD A), which was still being worked on at the completion of the second year of the contract, appears here. A complete description will be contained in the third year report.

The compiled word grammar is arranged by columns. Column 1 consists of all terms which are numbered 1, counting the terms in the rule from right to left; column 2 consists of all such number 2 right-side terms; etc. Column 0 is composed of all left-side terms and their associated dummy terms.

Each column is sorted by category symbol and subscript package. Associated with each unique category-symbol-and-subscript-package group is a list of all the rule numbers which contain this group. Column 0 is sorted by dummy terms as well as category symbol and subscript package. The rule numbers for all terminating right-side terms (counting from right to left) contain a pointer to the left-side term in column 0.

The compiled grammar is kept in a random disk file. Its directory, given a column number and a category symbol, points exactly to where all such groups containing that category symbol are found for that column. When the retrieval function, called F RNOS, is given a column number and category symbol, it returns a list of all rule numbers, the presence or absence of the four right side operators (B, E, F, and P), and the location of the associated subscript packages.

The workspace is also kept in a random disk file. Each file entry $n.m$ contains the following information:

- 1) a pointer to file entry $n.(m+1)$ if it exists,
- 2) which FE this is: FROM, TO, and NE (Number of Entry)
(note: $n.m \equiv TO.NE$),
- 3) pointers to all the FE's this rule builds upon,
- 4) the computed left-side resulting from antecedent construction,
- 5) the entire rule constructed from all the terms in the compiled grammar,
- 6) condition information.

The directory for the workspace points to the first and last FE for each file.

Both the compiled grammar and the workspace are loaded as the first step of WORD A. Then each existing file is processed from left to right.

The function PT1 (Process Term 1) is called first. It extracts the category symbol S of the FE and calls F RNOS with S and a column number of 1. If no rule exists for this combination, the next FE is tried. For each rule number found, the following instructions are carried out:

- 1) If an F-operator is present, check whether the file T0+1 has at least 1 FE with an M-condition. If it does not, discard this rule number.
- 2) Check whether this is a one-term rule. If not, go to 7).
- 3) If a P-operator is present, check whether the file FROM-1 has at least one FE with an M-condition. If it does not, discard this rule number.
- 4) Make an interim table entry in a table with all the relevant information.
- 5) Perform subscript check, value check, and operations. If the rule applies, perform antecedent (left-side) construction. If not, discard this interim entry and rule number.
- 6) Create a new FE, discard the interim table entry, and go to the next rule number.
- 7) Make a new table entry containing information about:
 - a) what FE has been matched by which category-subscript package, and,
 - b) what elements can match next, i.e., a column number (always 2), a file, and a rule number.

After all rule numbers are processed, the operation is repeated for the next FE until no more such FE's exist. At this stage, all degree 1 rules that can apply for this file have applied. All degree 2 or greater rules that have partially applied will have been converted into table entries. If there are no such table entries remaining, processing is complete for this file.

The three important items in the table referred to in 7b), above, are the Column, C, the File, F, and the Rule number, R. A function, SCAN T, is called, which computes the next file to be processed and all associated values of C. This next file chosen by SCAN T is the greatest file not exceeding the last file processed.

For each value of C a function PTX (Process Term n, where $n > 1$) is called. It is similar to PT1 described earlier. The main difference is that for each rule number retrieval by F RNOS a check is made against the table to determine whether the proper combination of C, F, and R are present in order for the new term to continue the rule. If there is a match, a new table

entry is constructed which contains all the previous information but with C increased by one and with a new value for F. The old entry is not modified, however, as it could still be built upon.

If the rule has ended, items 3) through 6) of PT1 are also performed, and a pointer to the new FE is placed in a list of newly created FE's .

After PTX has been performed for all the values of C for the next file, PT1 is executed for all the items in the newly-created FE list. In addition, any new FE's will also be placed on this list and subsequently processed.

This entire process is repeated for the next FE until the file is exhausted. SCAN T is then called again to pick the next file and column numbers to process. When this process finishes, control then returns to the main program where the entire process is repeated for the next available file.

2.2.3 Transfer (Normal Form) Grammar Sort

Using the transfer (normal form) grammar rules as input, the Transfer Grammar Sort program (TRAN GS) creates and sorts records consisting of—

- a) right-side terms of an explicit rule or an alpha switch rule
- b) any associated set of subscripts
- c) any alpha switch rule information pertaining to an explicit rule, and
- d) an associated left-side term.

The file consisting of these sorted records will be used as input to the Transfer Tree Construction program (TRAN TC).

The transfer (normal form) rules are coded by the linguists in a format representing a simple tree with right-pointers indicating continuation of rules and down-pointers indicating alternate rules with the same structure for nodes in $1-(n-i)$, where n is any node. The operators A and B in the input rule determine the shape of the new tree. The A operator indicates a node and the B operator indicates a specific branch to which a right-side term is to be attached.

When a right-side term in an input transfer rule (explicit rule) contains the operator A, a new pseudo rule (alpha switch rule) is created for this term and information indicating the creation of the new rule is added to the sort record for the explicit rule. One sort record is created for each input rule (explicit rule) and for each alpha switch rule.

TRAN GS invokes the CDC 6600 SORT/MERGE routine, and the resulting file (rearranged according to the A and B operators) is then used as input to TRAN TC.

2.3 Supporting Programs

A number of supporting programs were constructed in order to:

- a) update the English and German dictionaries (DICT UP)
- b) generate displays of the combined German verb and BOND lists
- c) convert the German-English noun list to subscript format
- d) convert the English verb list to subscript format
- e) convert the combined German verb and BOND lists to subscript format
- f) enable the workspace output by Dictionary Choice to be processed by the concordance program.

These are described in the following pages.

2.3.1 Dictionary Update

The dictionary update program (DICT UP) performs maintenance operations on files in dictionary-type format. It deletes, substitutes, or adds entries in a variety of ways as discussed below. With the exception of duplicate number changes, no modifications are made to the rule itself.

There are four possible inputs, any of which may be present or absent with the exception that Tape 2 and Tape 4 may not be present simultaneously. (If both are present, the file on Tape 2 will be replaced by sorted information from Tape 4.)

The types of input associated with each file are as follows:

Tape 1 contains the records to be updated. These must be in dictionary-type format and in rule-number/duplication-number order. If this file is not present, a new file is generated from the other input files.

Tape 2 contains update records applying to specific rule numbers. The entries must also be in dictionary format and rule-number/duplication-number order. Otherwise, they should appear on Tape 4 which sorts them automatically. These update records may perform the following operations—

- a) replace an existing record on the file to be updated (Tape 1) with the Tape 2 correction record having the same rule number.
- b) add a record with its associated rule number from the correction file (Tape 2) if there is no matching rule number in the file to be updated (Tape 1).
- c) delete a record from the file to be updated if the correction record has a rule of zero length.
- d) add a rule to the set of rules containing the same rule number but different duplication numbers by assigning either the first available duplication number, or the duplication number specified in the correction record. (If the existing record of the same rule number (Tape 1) has a duplication number of 0, it is changed to 1 and the correction record from Tape 2 is assigned the duplication number 2.)

Tape 3 contains additional records with no assigned rule numbers. Their format is somewhat different in that the rule number in the first word of the record is ignored and the rule number data set in the body of the rule is

omitted. This program inserts the newly assigned rule number into both places in the appropriate form.

The program assigns rule numbers to these rules whenever gaps occur in the file being created (Tape 5). Any remaining rules on the additional file (Tape 3) are added at the end with consecutive rule numbers. The duplication numbers are copied. A sequence of rules with a unique duplication number is assigned the same rule number, i.e., the same rule number is assigned to any sequence of numbers until a duplication number of 0 or 1 is encountered.

Tape 4 contains the same types of records as Tape 2. It is used when such records are not in rule-number/duplication-number order. The program sorts the records on this file, adds them to Tape 2, and proceeds as if the input had been Tape 2 initially.

The output consists of the updated dictionary on Tape 5, printed output displaying all updates made to the dictionary, and appropriate statistics. Tape 5 contains records in dictionary-type format.

The program is designed so that there are routines to perform all the update functions and a routine that decides which function is to be performed, depending on a) which files are empty, and b) the relationships among the values of rule numbers and duplication numbers of records in the three input files and the last record written on the updated file (Tape 5).

Initially the program determines whether the corrections file (Tape 4) is empty. If it is not, these records are sorted by rule-number/duplication-number to Tape 2. If Tape 4 is empty, corrections are taken from Tape 2. Then the first record of each file is read and the program performs updating operations as outlined in the input file description.

The organization of the program consists of a main logic program which handles the major part of the updating manipulations and several subroutines. The subroutines handle such functions as reading input records, writing new files of updated records, generating print output, page skipping of print output, and generating printable rule form.

The totals printed at the conclusion of the program, i.e., after all the input files are empty, provide the following statistics:

- a) number of old file records read from Tape 1
- b) number of correction records read from Tape 2
- c) number of addition records (without rule numbers) added from Tape 3

- d) number of records deleted from Tape 1
- e) number of old file records from Tape 1 to which the duplication number 1 was added
- f) number of records added from the correction file on Tape 2
- g) number of records replaced on Tape 1
- h) total number of output records written on Tape 5.

The program also produces error diagnostics for invalid operations and data, e.g., out of range rule number, out of range record length, attempt to delete non-existent record, and rule out of order.

2.3.2 Processing of the Bond List and the German Verb List

A number of programs having the name BOND n (where "n" is a number) were discussed in the monthly reports of the second year of the contract period. Four of these are described below.

BOND 1 compares the German Verb List and the Bond List and produces a merged German-Bond List and a display. Entries with corresponding request numbers in the original lists are combined into the German-Bond List. Any entry without an associated entry in both original files is displayed separately.

BOND 2 sorts the merged German-Bond List according to prefix-stem and displays the entire sorted list.

BOND 3 separates the German-Bond List into entries with a German part and only one English equivalent. The output of this program was used as input when these lists were subsequently converted to their combined subscript format.

BOND 4 sorts the separated German-Bond List (produced by BOND 3) in any order and in any of the four factorial possibilities based on the following items:

- G prefix-stem, German
- E English equivalent
- S subject information
- D area of provenience (domain).

Any of the four items may be omitted, resulting in the exclusion of all instances of the omitted item from the display. An additional feature of this program allows sorting of only those entries which contain items S, D, or both.

2.3.3 Conversion of the German-English Noun List to Subscript Format

Entries in the original German-English noun list (A through HA in Wildhagen [4]) were encoded as follows:

$N G; E \{ S E \}^*$

The referents of these symbols are—

N = German noun

G = gender information (i.e., $M F N PL$)

E = English equivalent encoded as $(P) EW O$

P = optional area of provenience (domain) information

EW = an English Phrase

O = optional object information consisting of a German preposition, GP , and an English preposition, EP , encoded as $[GP; EP]$

S = separator characters , and ;

$\{X\}^*$ = implies that X can occur any number of times.

Examples:

1. ABSCHIEBUNG F; DEPORTATION
2. ABONNENT M; SUBSCRIBER [AUF: TO]
3. ABORT M; (MED) ABORTION, MISCARRIAGE
4. ABREISSBLACK M; MEMORANDUM-BLOCK, TEAR-OFF BLOCK

The conversion of an entry gave each unit of information its own line as follows:

	line 1	= N word count
	line 2	= $GD(G)$
	line 11	= EW word count
The first E	line 12	= $\langle P \rangle$
	line 13	= $OP(GP) = OB(EP)$
	line 21	= EW word count
The second E	line 22	= $\langle P \rangle$
	line 23	= $OB(GP) = OB(EP)$

Word count consisted of the number of words in N or EW minus the number of hyphens.

Once an English equivalent contained area of provenience information, this same (P) with the addition of the asterisk was

assigned to all the following English equivalents which did not have a (P). (Cf. examples 3 above and below.)

After conversion, the entries given as examples above appear as follows—

1. GNO 468 1 ABSCHIEBUNG 1
GNO 468 2 GD(F)
GNO 468 11 DEPORTATION 1
2. GNO 406 1 ABONNENT 1
GNO 406 2 GD(M)
GNO 406 11 SUBSCRIBER 1
GNO 406 13 OB(AUF) = OB(TO)
3. GNO 410 1 ABORT 1
GNO 410 2 GD(M)
GNO 410 11 ABORTION 1
GNO 410 12 <MED>
GNO 410 21 MISCARRIAGE 1
GNO 410 22 <MED> *
4. GNO 434 1 ABREISSBLACK 1
GNO 434 2 GD(M)
GNO 434 11 MEMORANDUM - BLOCK 2-1
GNO 434 21 TEAR - OFF BLOCK 3-1

2.3.4 Conversion of the English Verb List (Hornby)

A program was designed to convert entries from the old English verb list to a subscripted form. Each entry in the verb list contained a main verb followed by one or more P-descriptors.

Each P-descriptor associated with a main verb was converted according to a format specified by the linguists. Thus, a new entry was created for each P-descriptor. Each new entry contained the main verb and a set of subscripts determined by the particular descriptor. These entries were then sorted, first according to the main verb and then according to the P-descriptor.

The request number for each occurrence of identical verbs was the same. The line numbers ranged from 001-008 to (n-1)1-(n-1)8, where n was the number of occurrences of a specific verb. For example:

<i>Request number</i>	<i>Line number</i>	<i>Entry</i>
NEVxxxx	001	main verb + word count for line 001
NEVxxxx	002	TY()
NEVxxxx	003	TS()
NEVxxxx	004	FS()
NEVxxxx	005	TO()
NEVxxxx	006	OB()
NEVxxxx	007	RA()
NEVxxxx	008	P-descriptor used in conversion

An example of the form of an original English entry:

EVLO004089 FIDGIT VI VT P21 PIH
EVLO019234 FIDGIT WITH P24K

The form of the converted entry is:

NEV1852001 FIDGIT
NEV1852002 TY(VT,VI)
NEV1852005 TO(HU)
NEV1852008 P21 PIH
NEV1852011 FIDGIT 1
NEV1852012 TY(VT)
NEV1852015 TO(IN,AB)
NEV1852016 OB(WITH)
NEV1852018 P24K

2.3.5 Conversion of the Bond List and the German Verb Lists to Subscript Format

A program was designed to convert simultaneously the German verb list and its English translation equivalents (the BOND list) to subscripted form. The input to this program consisted of the German verb list merged by request numbers with the BOND list.

Each German entry was converted according to a set of rules supplied by the linguists. Next, each BOND entry or entries for a specific German entry was converted according to the descriptor patterns within each English entry. The entire list was then sorted according to the German verbs, using a prefix-stem sort. The BOND entries within each German entry were sorted alphabetically by line 000 of the converted list.

The program produced a combined German-BOND list in subscripted form. The request number for a German entry and all its associated English entries was the same. The line numbers increased sequentially within a complete entry.

For example,

<u>Request No.</u>	<u>Line No.</u>	<u>Entry</u>
GBDxxxx	000	[PREFIX] Verb + no. words line 000 - no. words in main verb
GBDxxxx	001	TY()
GBDxxxx	002	FS()
GBDxxxx	003	TS()
GBDxxxx	004	DS()
GBDxxxx	005	OB()
GBDxxxx	006	TO()
GBDxxxx	007	RA()
GBDxxxx	008	OA()
GBDxxxx	009	AUX()
GBDxxxx	010	Entire old German entry (in- cluding old request no.)
GBDxxxx	*n0	Main verb [ADPREP] [NOISE] + no. of words in n0 counting only the Main verb and the ADPREPS
GBDxxxx	n1	TY()
GBDxxxx	n2	FS()

<u>Request No.</u>	<u>Line No.</u>	<u>Entry</u>
GBDxxxx	n3	TS()
GBDxxxx	n5	OB()
GBDxxxx	n6	TO()
GBDxxxx	n7	RA()
GBDxxxx	n8	OA()
GBDxxxx	n10	Entry from line 010
GBDxxxx	(n+1)0	Entry from line 000
GBDxxxx	(n+1)1	TY() from line 001
GBDxxxx	(n+1)2	FS() from line 002
GBDxxxx	(n+1)3	TS() from line 003
GBDxxxx	(n+1)5	OB() from line 005
GBDxxxx	(n+1)6	TO() from line 006
GBDxxxx	(n+1)7	< > selectors
GBDxxxx	(n+1)8	() selectors
GBDxxxx	(n+1)9	The entire English entry

*n = 2J, J=1,2,...,m where m = number of English entries
associated with a given
German entry

The original German entry
GRW0314020 AUS WASCH VT AUS
with the associated BOND portion
(GD SCHMUTZ) WASH OUT OF
now have the following form as a converted entry:

GBD2798001 TY(VT)
GBD2798005 OB(A + AUS)
GBD2798006 TO(* + *)
GBD2798009 AUX(H)
GBD2798010 GRW0314020 AUS WASCH VT AUS
GBD2798020 WASH I
GBD2798021 TY(VT)
GBD2798025 OB(OUT OF)
GBD2798026 TO(*)
GBD2798029 GRW0314020 AUS WASCH VT AUS

GBD2798030 AUS WASCH 2-1
GBD2798031 TY(VT)
GBD2798035 OB(A + AUS)
GBD2798036 TO(* + *)
GBD2798038 GD SCHMUTZ
GBD2798039 (GD SCHMUTZ) WASH OUT OF

2.3.6 Concordance of the Dictionary Choice Workspace

Two programs were constructed to enable the workspace output from Dictionary Choice to be processed by the concordance program in order to gain statistics on situations involving compounding and ending rules. Below is a brief, very general explanation of what the programs do, followed by an extensive example.

The Program CONDCOP takes as input the workspace output by Dictionary Choice and builds three lists from it as follows:

- a) the character strings of all words which had only one rule spanning them. This list was designed as text that could be re-analyzed by the dictionary and choice programs.
- b) all words which had more than one rule spanning them. This list contains a separate entry for each unique span of the word in question. The information contained in each entry includes the character string of the entire word and the left-side of the rule that analyzed it, and the character string of each subspan and its associated left-side. This list was designed to be processed by the concordance program.
- c) the left-side terms of List b) in two different formats. This list was designed to be processed by the glossary program.

The second program, CONVRT 1, takes List b) above as input and preprocesses it for the concordance program by a set of rules. The output is a new list utilizing the non-standard processing mode of the concordance program where only those items desired to be concorded are tagged accordingly.

EXAMPLE

Take as text the 22-character string:

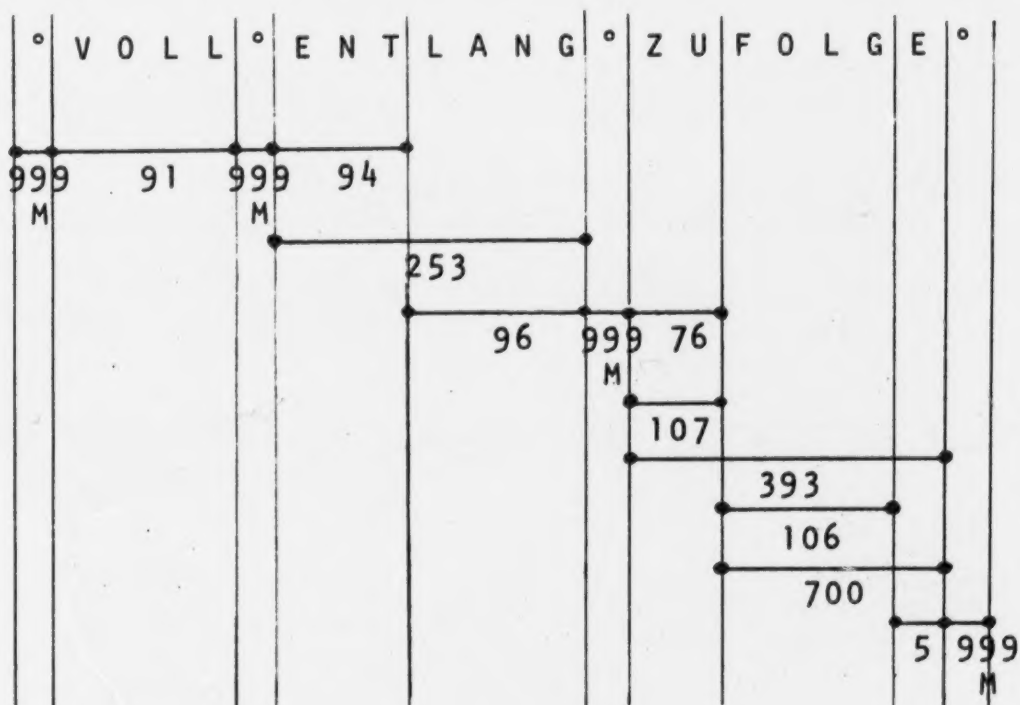
°VOLL°ENTLANG°ZUFOLGE°

Process the text by Dictionary Analysis and Choice (K-option off) with the rules:

C 5	V ENDG	* E
	+ TY(E)	B
C 76	V ENDG	* ZU
	+ TY(ZU)	P

C 91	V A + CV + IN + PV(O) + SV(S)	* VOLL
C 94	V PRFX	* ENT P
C 96	V A	* LANG
C 106	V V	* FOLG
C 107	V PREP	* ZU P
C 253	V PREP + GC(D) + POS(POST)	* ENTLANG P
C 393	V PREP + GC(D) + POS(POST)	* ZUFOLGE P
C 700	V N + CL(34) + GD(F)	* FOLGE
C 999	V SPACE M	* E

This results in the simulated workspace diagrammed below:



CONDCOP processes the three words (sequence of characters between M-symbols) in the workspace and generates as output the three lists as follows:

List (1)

DC1 1 VOLL

List (2)

DC2 1 .ENTLANG. ENT LANG -PRFX -A *PREP:GC(D):POS(POST)*

DC2 2 .ZUFOLGE. ZU FOLG E -ENDG:TY(ZU) -V -ENDG:TY(E)
PREP:GC(D):POS(POST)

DC2 3 .ZUFOLGE. ZU FOLGE -ENDG:TY(ZU) -N:CL(34):GD(F)
PREP:GC(D):POS(POST)

DC2 4 .ZUFOLGE. ZU FOLG E -PREP -V -ENDG:TY(E)
PREP:GC(D):POS(POST)

DC2 5 .ZUFOLGE. ZU FOLGE -PREP -N:CL(34):GD(F)
PREP:GC(D):POS(POST)

List (3)

DC3 1 PRFX-A*PREP:GC(D):POS(POST)*
*PREP:GC(D):POS(POST)*PRFX-A

DC3 2 ENDG:TY(ZU)-V-ENDG:TY(E)*PREP:GC(D):POS(POST)*
*PREP:GC(D):POS(POST)*ENDG:TY(ZU)-V-ENDG:TY(E)

DC3 3 ENDG:TY(ZU)-N:CL(34):GD(F)*PREP:GC(D):POS(POST)*
*PREP:GC(D):POS(POST)*ENDG:TY(ZU)-N:CL(34):GD(F)

DC3 4 PREP-V-ENDG:TY(E)*PREP:GC(D):POS(POST)*
*PREP:GC(D):POS(POST)*PREP-V-ENDG:TY(E)

DC3 5 PREP-N:CL(34):GD(F)*PREP:GC(D):POS(POST)*
*PREP:GC(D):POS(POST)*PREP-N:CL(34):GD(F)

The four entries for ZUFOLGE come from the four unique spans:

- 1) 76 - 106 - 5
- 2) 76 - 700
- 3) 107 - 106 - 5
- 4) 107 - 700

CONVRT 1 processes List b) data and adds the character "+" in positions that can be calculated by a set of rules to be defined below. Also, any unit of data that does not generate at least one "+" will be eliminated from the output.

The rules are divided into two classes: one or more arithmetic assignment statements and one logical function.

The arithmetic assignment statements are used to define variables that will be used in the logical function. They have the form:

$$A = \{ - \mid * \}^1 \{ CS \}^1 \{ \{ :SN \}^1 \{ / \}^1 \{ (\{ BO \mid UO \mid SV \mid ^\circ \}^*) \}^1 \}^*$$

where A = any letter of the alphabet

CS = category symbol

SN = subscript name

UO = unary operator

BO = binary operator

SV = subscript value

$^\circ$ = blank space

$\{ X \}^1$ = X may occur either zero or one time only

$\{ X \}^*$ = X may occur any number of times including zero

$\{ X \mid Y \}$ = either X or Y

The arithmetic assignment statements in effect define skeletons of the left-side terms of the List b) data which will be used for comparison purposes in conjunction with the logical function.

The logical function G can be defined by the BNF grammar:

$G = * F$

$F = A$

$F = F , A$

$F = F + A$

$F = (F)$

where A is a variable defined by an arithmetic assignment statement. Comma stands for *or* and plus stands for *and*.

Examples of rules CONVRT 1 might use to preprocess the
List b) data:

1) A=:GC
*A

Concord those entries which have a subscript name (SN)
of "GC" on the SN itself.

2) A=-N
B=-V
C=-A
*A,B,C

Concord on the "-" followed directly by the category symbol
(CS) those entries which have at least one subspan having
a CS of "N", "V", or "A".

3) A=-ENDG:TY(E)
B=-ENDG:TY(ZU)
*A,B

Concord on the "-" followed directly by the CS those en-
tries containing a subspan having a CS of "ENDG" with a
SN of "TY" with either a subscript value (SV) "E" or "ZU".

4) A=-N
B=-V
C=*PREP
*(A,B)+C

Concord on the "-" followed directly by the CS and on the
"*" followed directly by the CS those entries containing
a subspan having a CS of either "N" or "V" with a complete
span having a CS of "PREP". In other words, all cases
where a noun or a verb (in conjunction with any number of
other terms) has been rewritten as a preposition.

5) A=(D)
*A

Concord on the SV those entries having a SV of "D".

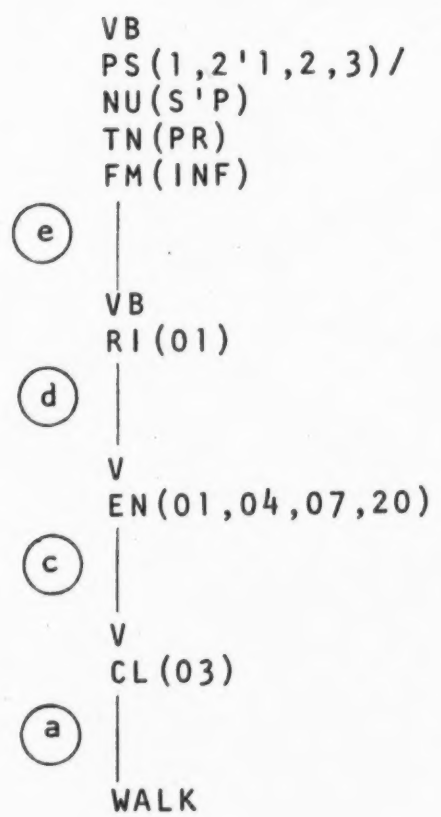
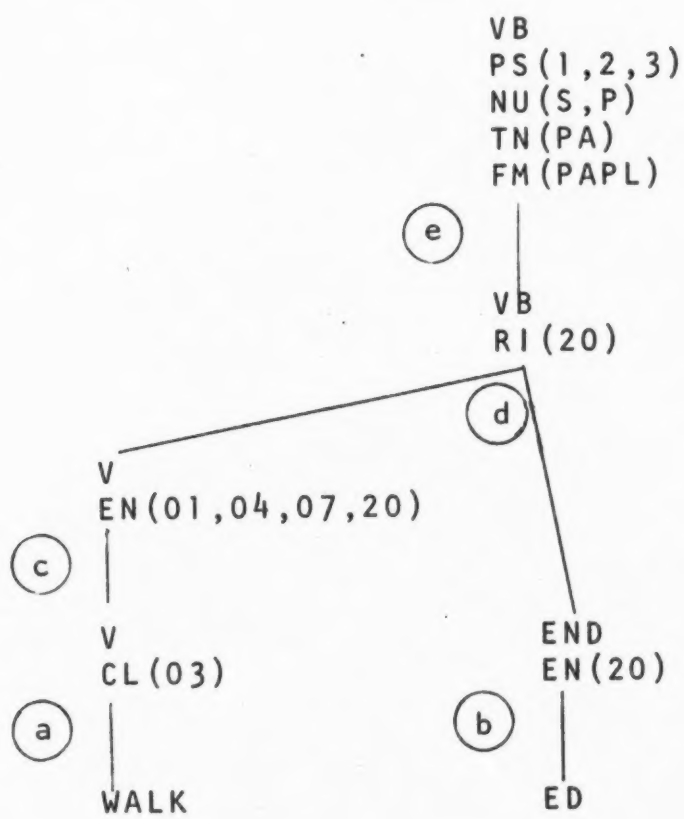
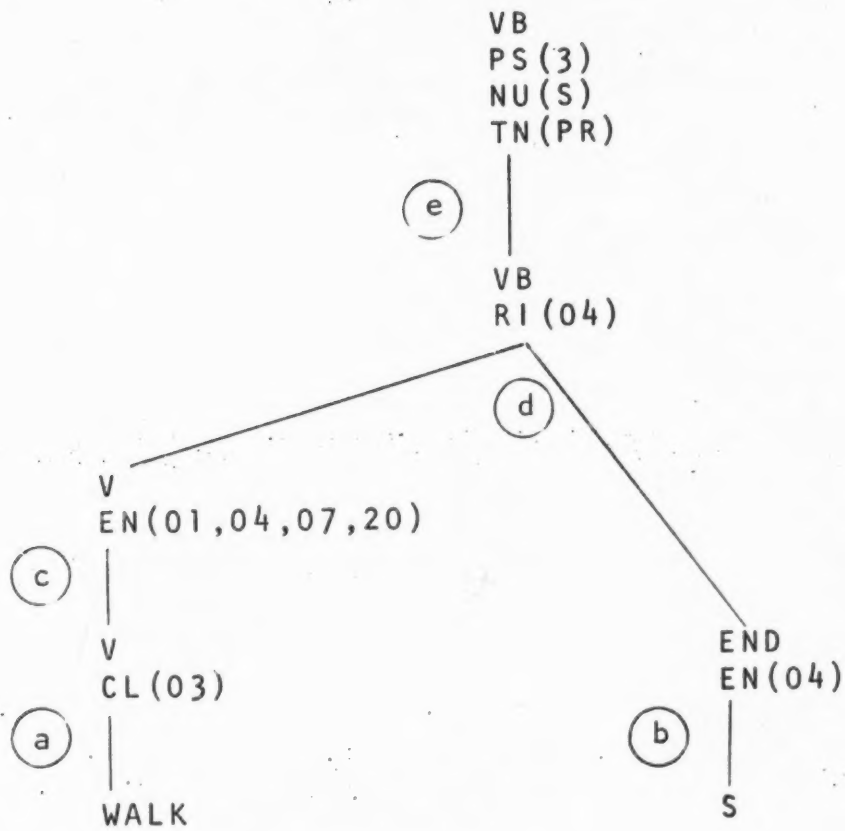
CONCLUSION

Linguistic work during the period was directed at expanding the dictionaries for both German and English. The number of lexical items coded was increased, as were the features noted for each item. The coding involved linguists in some of the most complex problems of linguistic description, such as the treatment of adverbs and extended forms of verbs. Adverbs have unfortunately not been adequately classified in any of the grammars, and accordingly a new system of classification has had to be developed. Fortunately the extended forms of German verbs have been treated in a recent monograph, which was used as a start for our own lexical analysis. These extended forms are particularly difficult in having a small group of verbs used primarily to depict the verbal component and nouns combined with them to carry the semantic portion of the expression.

Like the linguistic analysis, programming is proceeding according to schedule. Imaginative programs have enabled us to identify the specific problems encountered in our texts. Others make up the massive set of programs necessary to manage the deep as well as the surface structures of language. If the current momentum can be maintained, we should be capable of carrying out on schedule the projected demonstration of German-to-English translation.

APPENDIX A

VERB MORPHOLOGY



- a) = dictionary classification of verb
 b) = identification of ending
 c) = redundancy rule which indicates the set of endings the verb may concatenate with
 d) = stem-ending concatenation rule
 e) = redundancy rule specifying the grammatical features of the verb form
- Set (e) is not included in the word grammar but is part of the syntax.

(a) Dictionary Verb Classes:

S/C Class	Base	3 Sing.	Present Partic.	Past	Past Partic.	Example
01	E	ES	ING	ED	ED	revolv
02	+	ES	ING	ED	ED	reach miss
03	+	S	ING	ED	ED	soar
04	+	S	BING	BED	BED	rub
05	+	S	DING	DED	DED	stud
06	+	S	GING	GED	GED	plug
07	+	S	KING	KED	KED	panic frolic
08	+	S	LING	LED	LED	pal
09	+	S	MING	MED	MED	stem
10	+	S	NING	NED	NED	don
11	+	S	PING	PED	PED	stop
12	+	S	RING	RED	RED	blur
13	+	SES	SING	SED	SED	gas
14	+	S	TING	TED	TED	knit
15	+	ZES	ZING	ZED	ZED	quiz
16	+	S	ING	D	D	agree
17	+	S	ING	ED	N	show
18	+	S	ING	+	+	read
19	+	S	ING			dream
20	+	ES	ING			focus
21	E	ES	ING			mak

S/C Class	Base	3 Sing.	Present Partic.	Past	Past Partic.	Example
22				+	N	wove
23				E	N	swor
24				+	+	unwound
25	+	S	ING		N	see
26	+	S	ING		EN	eat
27	E	ES	ING		EN	giv
28				+		fell ate
29	+	S		+	+	cut
30			ING			cutt
31	+		ING			cry imply
32		ES		ED	ED	cri impli
33					+	lain

(b) Dictionary Rules for V Endings:

V END
+ EN(02,34)
D
B

= * E

V END
+ EN(07)
D
B

= * ING

V END
+ EN(03)
D
B

= * ES

V END
+ EN(08)
D
B

= * BING

V END
+ EN(04)
D
B

= * S

V END
+ EN(09)
D
B

= * DING

V END
+ EN(05)
D
B

= * SES

V END
+ EN(10)
D
B

= * GING

V END
+ EN(06)
D
B

= * ZES

V END
+ EN(11)
D
B

= * KING

V END = * LING
+ EN(12)
D
B

V END = * MING
+ EN(13)
D
B

V END = * NING
+ EN(14)
D
B

V END = * PING
+ EN(15)
D
B

V END = * RING
+ EN(16)
D
B

V END = * SING
+ EN(17)
D
B

V END = * TING
+ EN(18)
D
B

V END = * ZING
+ EN(19)
D
B

V END = * ED
+ EN(20)
D
B

V END = * BED
+ EN(21)
D
B

V END = * DED
+ EN(22)
D
B

V END = * GED
+ EN(23)
D
B

V END = * KED
+ EN(24)
D
B

V END = * LED
+ EN(25)
D
B

V END = * MED
+ EN(26)
D
B

V END = * NED
+ EN(27)
D
B

V END = * PED
+ EN(28)
D
B

V END = * RED
+ EN(29)
D
B

V END = * SED
+ EN(30)
D
B

V END = * TED
+ EN(31)
D
B

V END = * ZED
 + EN(32)
 D
 B

V END = * D
 + EN(33)
 D
 B

V END = * N
 + EN(37)
 D
 B

V END = * EN
 + EN(38)
 D
 B

Values of the S/C EN (ending of verbs):

01	=	Ø as base ending
02	=	E "
03	=	ES as 3 sg. ending
04	=	S "
05	=	SES "
06	=	ZES "
07	=	ING as prpl. ending
08	=	BING "
09	=	DING "
10	=	GING "
11	=	KING "
12	=	LING "
13	=	MING "
14	=	NING "
15	=	PING "
16	=	RING "
17	=	SING "
18	=	TING "
19	=	ZING "
20	=	ED as past and papl. ending *
21	=	BED "
22	=	DED "
23	=	GED "
24	=	KED "
25	=	LED "
26	=	MED "
27	=	NED "
28	=	PED "
29	=	RED "
30	=	SED "
31	=	TED "
32	=	ZED "
33	=	D "
34	=	E as past ending
35	=	Ø as past ending
36	=	Ø as papl. ending
37	=	N "
38	=	EN "

* In the verb class CL(17), ED forms only the past.

In the redundancy rule
$$\begin{array}{ccc} V & = & V \\ \text{EN}(\dots 20) & & \text{CL}(17) \end{array}$$

which defines the endings that go with class 17, the S/C NP (not papl) is added.

In the redundancy rule
$$\begin{array}{ccc} \text{VB} & = & \text{VB} \\ \text{PS} & & \text{RI} \\ \text{NU} & & \\ \text{TN} & & \end{array}$$

which defines the person, number, etc. of the full verb, the S/C FM(PAPL) is introduced only if the S/C NP is not present in the VB label.

(C) Redundancy Rules Dominating V:

$$\begin{array}{ccc} V & V & \\ + \text{EN}(02,03,07,20) & = & V & V \\ \wedge 2 & & \$ & \text{CL}(01) \end{array}$$

$$\begin{array}{ccc} V & V & \\ + \text{EN}(01,03,07,20) & = & V & V \\ \wedge 2 & & \$ & \text{CL}(02) \end{array}$$

$$\begin{array}{ccc} V & V & \\ + \text{EN}(01,04,07,20) & = & V & V \\ \wedge 2 & & \$ & \text{CL}(03) \end{array}$$

$$\begin{array}{ccc} V & V & \\ + \text{EN}(01,04,08,21) & = & V & V \\ \wedge 2 & & \$ & \text{CL}(04) \end{array}$$

$$\begin{array}{ccc} V & V & \\ + \text{EN}(01,04,09,22) & = & V & V \\ \wedge 2 & & \$ & \text{CL}(05) \end{array}$$

.

.

.

.

$$\begin{array}{ccc} V & V & \\ + \text{EN}(01,04,07,20,37) & = & V & V \\ + \text{NP} & & \$ & \text{CL}(17) \\ \wedge 2 & & \end{array}$$

.

.

.

.

(d) Verb plus Ending Concatenation Rules:

V VB	=	V V	V END
+*3.1RI		\$ EN	. 2.1EN
Λ 2			B
D 3			

V VB	=	V V
\$ 2.1RI		\$ EN(01,35,36)
Λ 2		F

(e) Redundancy Rules Dominating VB:

V VB	=	V VB
+ PS(1,2'1,2,3)/		\$ RI(01,02)
+ NU(S'P)		
+ TN(PR)		
+ FM(INF)		
Λ 2		

V VB	=	V VB
+ PS(3)		\$ RI(03,04,05,06)
+ NU(S)		
+ TN(PR)		
Λ 2		

V VB	=	V VB
+ FM(PRPL)		\$ RI(07,08,09,10,11,12,13,
Λ 2		14,15,16,17,18,19)

V VB	=	V VB
+ PS(1,2,3)		\$ RI(20,21,22,23,24,25,26,
+ NU(S,P)		27,28,29,30,31,32,33)
+ TN(PA)		* NP
+ FM(PAPL)		
Λ 2		

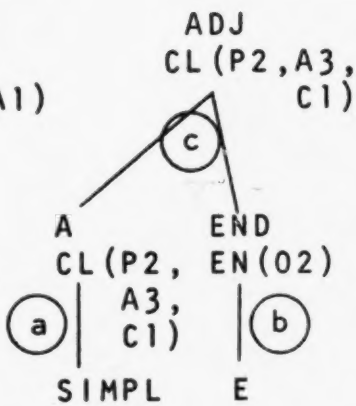
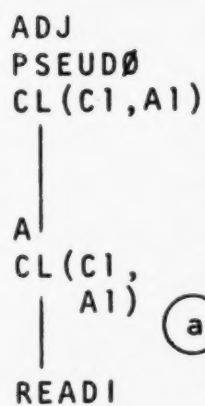
V VB	=	V VB
+ PS(1,2,3)		\$ NP
+ NU(S,P)		\$ RI(20)
+ TN(PA)		
Λ 2		

V VB	=	V VB
+ PS(1,2,3)		\$ RI(34,35)
+ NU(S,P)		
+ TN(PA)		
Λ 2		

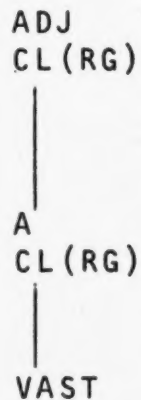
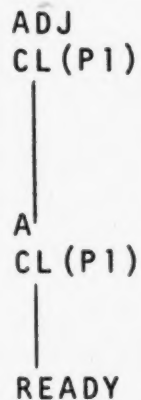
V VB	=	V VB
+ FM(PAPL)		\$ RI(36,37,38)
Λ 2		

ADJECTIVE MORPHOLOGY

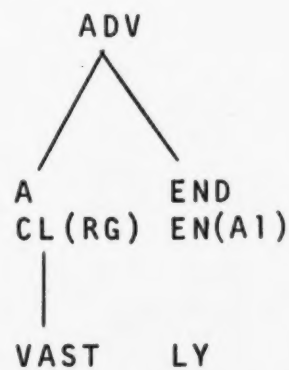
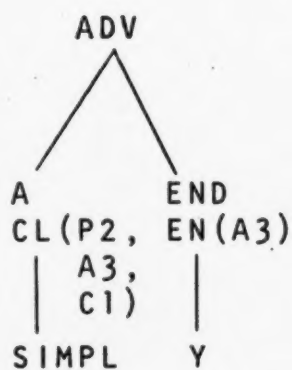
Pseudo-Positives



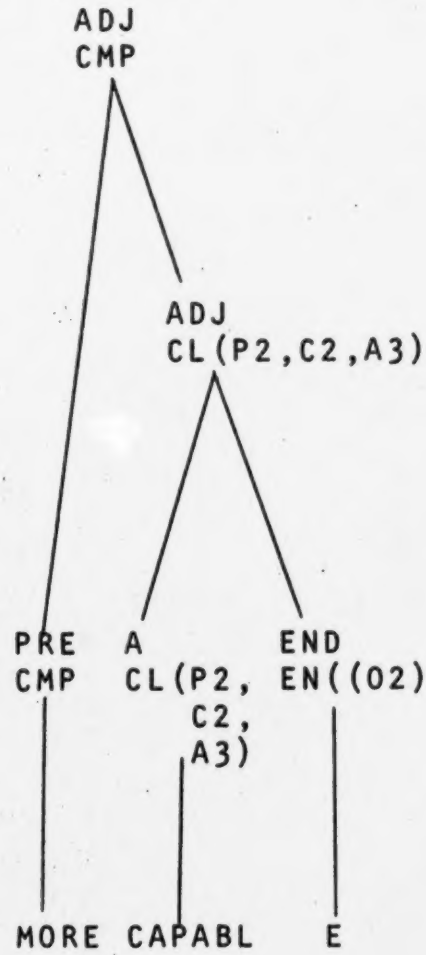
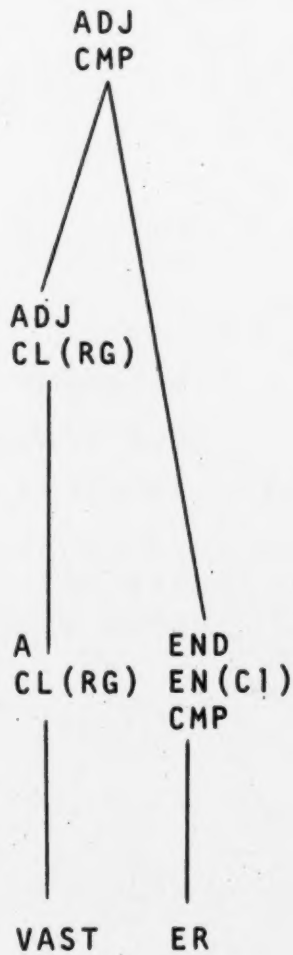
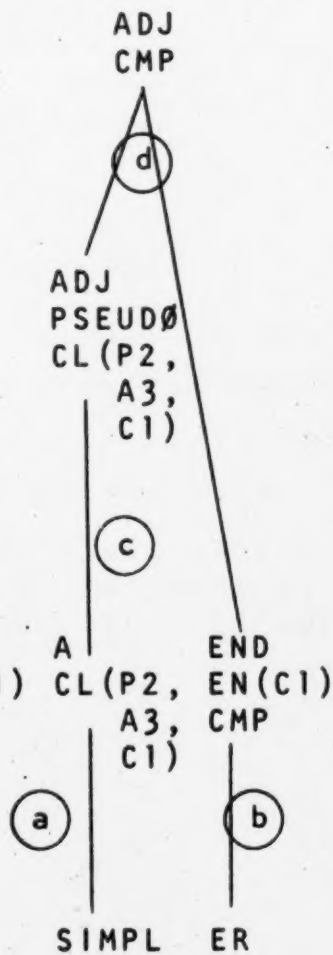
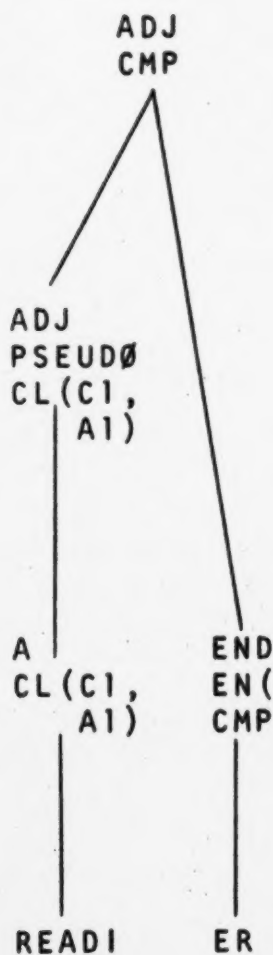
Positives



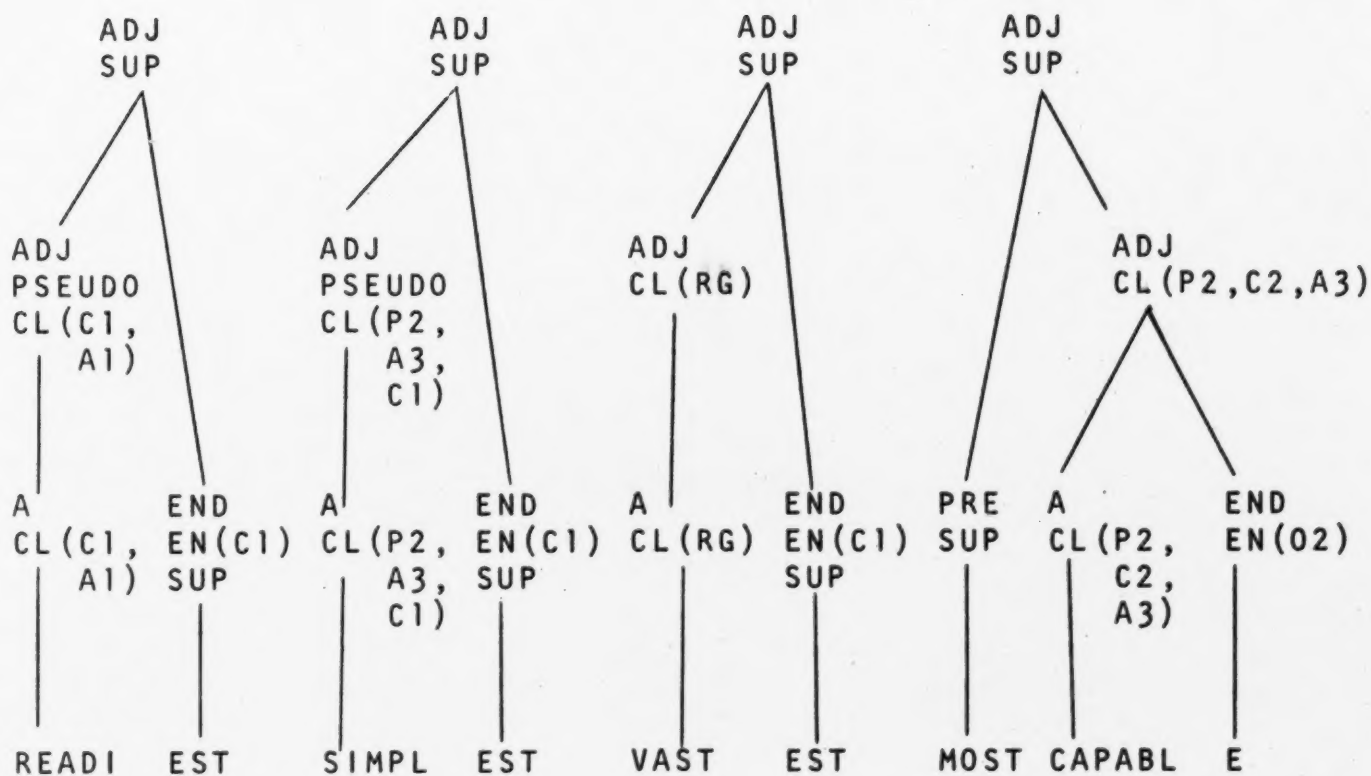
Adverbs



Comparatives



Superlatives



ADJ
SUP
|
BEST

ADJ
SUP
|
LEAST

- a) = dictionary classification of adjective
- b) = identification of ending
- c) = stem-ending concatenation rule
- d) = rule concatenating positive (or pseudo-positive) with comparative and superlative endings or *more* and *most*

(a) Adjective Dictionary Classes:

Webster	Class	Pos.	Comp.	+ Sup.	Adv.	Example:
A-A	CL(RR*)	+	more	most	ly	lateral
A-B	CL(P1,C2)	+	more	most		Greek
A-C	CL(P1,C2,A2)	+	more	most	ally	photographic
A-D	CL(P1)	+				ready
A-E	CL(P2,C2,A3)	e	more	most	y	capabl
A-F	CL(P1,C2,A4)	+	more	most	+	baby-like
A-G	CL(C1,A1)		er	est	ly	readi
A-H	CL(A1)				ly	benedictori
A-J	CL(RG*)	+	er	est	ly	vast
A-L	CL(P2,C1,A5)	e	er	est	ely	clos
A-M	CL(P1,A4)	+			+	far
A-N	CL(P2,C1,A3)	e	er	est	y	simpl, abl
A-P	CL(P1,C1)	+	er	est		few
A-Q	CL(P1,C1,A4)	+	er	est	+	low
A-R	CL(P1,C1,A3)	+	er	est	y	full
A-S	CL(P2,A3)	e			y	singl
A-T	CL(P2,C1)	e	er	est		whit
<u>RMD</u>						
A010(x)	CL(P1,C3,A1)	+	ger	gest	ly	snug
A011(x)	CL(P1,C4,A1)	+	mer	mest	ly	dim
A012(x)	CL(P1,C5,A1)	+	ner	nest	ly	thin
A013(x)	CL(P1,C6,A1)	+	ber	best	ly	glib, drab
A014(x)	CL(P1,C7,A1)	+	ter	test	ly	hot
A015(x)	CL(P1,C8)	+	der	dest		red
A016(x)	CL(P1,C5)	+	ner	nest		tan
A017(x)	CL(P1,C7)	+	ter	test		fat
A018(x)	CL(P1,C3)	+	ger	gest		big
A019(x)	CL(P1,C4)	+	mer	mest		trim

*regular, Germanic

Onset: Vowel onset is marked by the S/C ON without values;
consonant onset is unmarked.

(b) Adjective Endings:

P1 = ∅ for positive
A4 = ∅ for adverb

(V-ending)

V END = * E
+ EN(02,34)
D

V END = * LY
+ EN(A1)
+ AV
D

V END = * ALLY
+ EN(A2)
+ AV
D

V END = * Y
+ EN(A3)
+ AV
D

V END = * ELY
+ EN(A5)
+ AV
D

V END = * ER
+ EN(C1)
+ CMP
D

V END = * EST
+ EN(C1)
+ SUP
D

V END = * GER
+ EN(C3)
+ CMP
D

V END = * GEST
+ EN(C3)
+ SUP
D

V END = * MER
+ EN(C4)
+ CMP
D

V END = * MEST
+ EN(C4)
+ SUP
D

V END = * NER
+ EN(C5)
+ CMP
D

V END = * NEST
+ EN(C5)
+ SUP
D

V END = * BER
+ EN(C6)
+ CMP
D

V END = * BEST
+ EN(C6)
+ SUP
D

V END = * TER
+ EN(C7)
+ CMP
D

V END = * TEST
+ EN(C7)
+ SUP
D

V END = * DER
+ EN(C8)
+ CMP
D

V END = * DEST
+ EN(C8)
+ SUP
D

V PRE = * MØRE
+ CMP

V PRE = * MØST
+ SUP

<u>ADV</u>	<u>PØS</u>	<u>CØMP</u>	<u>SUP</u>
A1 LY	P1 Ø	C1 ER	EST
A2 ALLY	P2 E	C2 MORE	MOST
A3 Y		C3 GER	GEST
A4 Ø		C4 MER	MEST
A5 ELY		C5 NER	NEST
		C6 BER	BEST
		C7 TER	TEST
		C8 DER	DEST

(c) Adjective Concatenation Rules:

positive	V ADJ \$ 2.1 Λ 2	= V A \$ CL(RR, RG, P1)	
	V ADJ . 2.1(-C1) Λ 2 D 3	= V A \$ CL(P2)	V END \$ EN(02) B
pseudo- positive	V ADJ + PSEUDØ \$ 2.1 Λ 2	= V A \$ CL(P2+C1, C1+A1)	
	V ADV \$ 2.1	= V A \$ CL(A4)	
adverb	V ADV D 3	= V A \$ CL	V END . 2.1EN \$ AV B

(d)

comp. and superl.	V ADJ Λ 3 Λ 2 D 3	= V ADJ \$ CL	V END . 2.1EN * AV B
	V ADJ Λ 3 Λ 2 D 3	= V ADJ \$ CL(RG)	V END \$ EN(C1) B

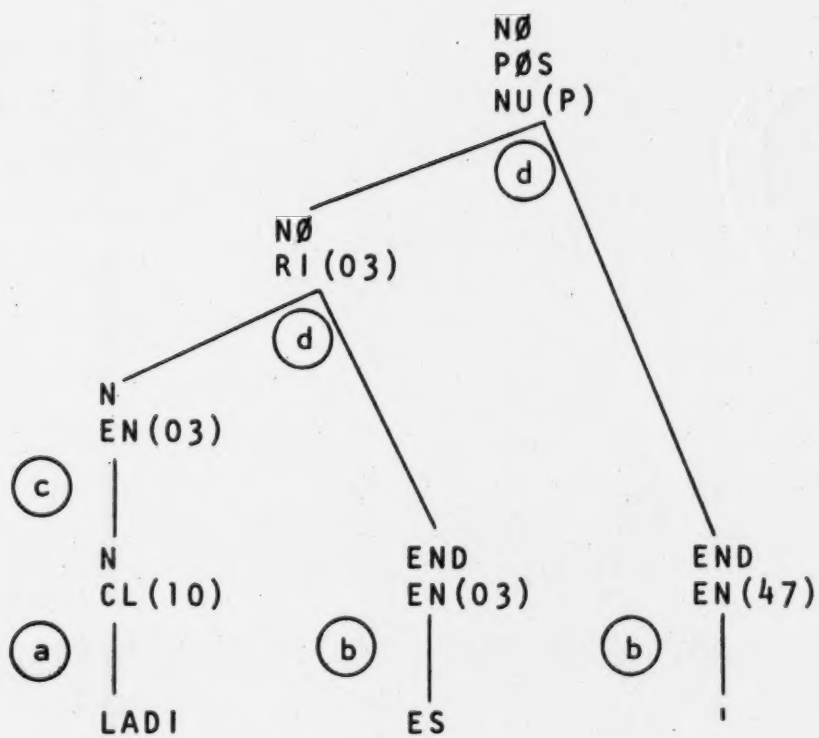
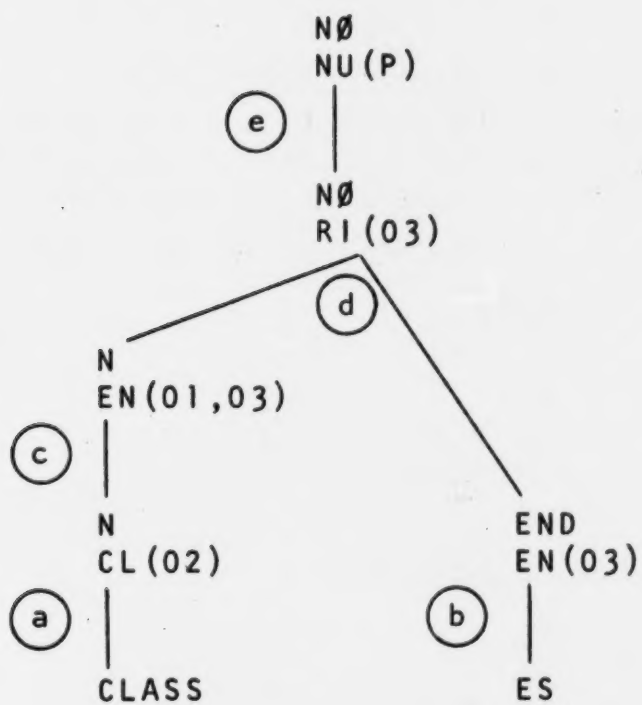
$$\begin{array}{l} V \text{ ADJ} \\ \wedge 2 \\ \wedge 3 \end{array} = V \text{ PRE} \quad V \text{ ADJ} \\ \$ \text{ CL}(C2,RR)$$

The label ADJ is excluded from any syntactic concatenation rule.
 PSEUDØ

Word Rules

V WØRD D 1,2,4	=	V LB	V NO B	V RB B
V WØRD D 1,2,4	=	V LB	V ADJ B	V RB B
V WØRD D 1,2,4	=	V LB	V VB B	V RB B
V WORD D 1,2,4	=	V LB	V ADV B	V RB B

NOUN MORPHOLOGY



- a) = dictionary classification of noun
- b) = identification of ending
- c) = redundancy rule indicating the set of endings the noun may occur with
- d) = stem-ending concatenation rule
- e) = redundancy rule specifying the grammatical features of the inflected noun. Set e) is not included in the word grammar but is part of the syntax.

onset: vowel onset is marked in the dictionary:

V N = * ALTAR
 + CL(01)
 + TY(N)
 + ØN

Consonant onset is unmarked.

(a) Noun Classes:

<u>Class</u>	<u>SG</u>	<u>Pl</u>	<u>Poss</u> (if unpredict- table from inflectional endings)	<u>Ex:</u>
01	+	S		work, altar
02	+	ES		apparatus, class
03	E	ES		chang(e)
04	+		'S	study, intensity
05	+		'	optics
06	+	+	'S	sheep, aircraft
07	+	+	'	series
08		+	'S	men
09		+	'	groats
10		ES		studi
11	+	E		stria, alumna
12	+	TA		stroma
13	IS	ES		cris, analys
14	ON	A		criteri, automat
15	UM	A		dat
16	US	I		radi
17	+	'S		A, B, C . . .

(b) Dictionary Rules for Noun Endings:

(V ending)	V END = *E + EN(02,34) D	01 = Ø for sg
	V END = *IS + EN(39) D	07 = Ø for pl
	V END = *ON + EN(40) D	02 = e for sg
	V END = *UM + EN(41) D	34 = e for pl
	V END = *US + EN(42) D	43 = 's
	V END = *'S + EN(43) D	47 = '
(V ending)	V END = *S + EN(04) D	
(V ending)	V END = *ES + EN(03) D	
	V END = *TA + EN(44) D	
	V END = *A + EN(45) D	
	V END = *I + EN(46) D	
	V END = *' + EN(47) D	

(c) Redundancy Rules Dominating N:

$$\begin{array}{l} V N \\ + EN(01,04) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(01) \end{array}$$

$$\begin{array}{l} V N \\ + EN(01,03) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(02) \end{array}$$

$$\begin{array}{l} V N \\ + EN(02,03) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(03) \end{array}$$

$$\begin{array}{l} V N \\ + EN(01) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(04) \end{array}$$

$$\begin{array}{l} V N \\ + EN(01) \\ + FS \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(05) \end{array}$$

$$\begin{array}{l} V N \\ + EN(01,07) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(06) \end{array}$$

$$\begin{array}{l} V N \\ + EN(01,07) \\ + FS \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(07) \end{array}$$

$$\begin{array}{l} V N \\ + EN(07) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(08) \end{array}$$

$$\begin{array}{l} V N \\ + EN(07) \\ + FS \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(09) \end{array}$$

$$\begin{array}{l} V N \\ + EN(03) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(10) \end{array}$$

$$\begin{array}{l} V N \\ + EN(01,34) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(11) \end{array}$$

$$\begin{array}{l} V N \\ + EN(01,44) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(12) \end{array}$$

$$\begin{array}{l} V N \\ + EN(39,03) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(13) \end{array}$$

$$\begin{array}{l} V N \\ + EN(40,45) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(14) \end{array}$$

$$\begin{array}{l} V N \\ + EN(41,45) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(15) \end{array}$$

$$\begin{array}{l} V N \\ + EN(42,46) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ CL(16) \end{array}$$

$$\begin{array}{l} V N \\ + EN(01,43) \\ \wedge 2 \end{array} = \begin{array}{l} V N \\ \$ CL(17) \end{array}$$

(d) Noun Concatenation Rules:

Singular and Plural:

V NØ	=	V N	V END
\$*3.1RI		\$ EN	. 2.1EN
Λ 2			B
D 3			

V NØ	=	V N
\$ 2.1RI		\$ EN(01,07)
Λ 2		

Possessives:

V NØ	=	V NØ	V END
+ PØS		\$ RI(39,42)	\$ EN(47)
+ NU(S)			B
Λ 2			
D 3			

V NØ	=	V NØ	V END
+ PØS		\$ RI(04,03)	\$ EN(47)
+ NU(P)			B
Λ 2			
D 3			

V NØ	=	V NØ	V END
+ PØS		\$ RI(01,02,	\$ EN(43)
+ NU(S)		40,41)	B
Λ 2		* FS	
D 3			

V NØ	=	V NØ	V END
+ PØS		\$ RI(07,34,	\$ EN(43)
+ NU(P)		44,45,46)	B
Λ 2		* FS	
D 3			

V NØ	=	V NØ	V END
+ PØS		\$ RI(01)	\$ EN(47)
+ NU(S)		\$ FS	B
Λ 2			
D 3			

V NØ	=	V NØ	V END
+ PØS		\$ RI(07)	\$ EN(47)
+ NU(P)		\$ FS	B
Λ 2			
D 3			

(e) Redundancy Rules Dominating NØ:

$$\begin{array}{l} V \text{ NØ} \\ + \text{ NU(S)} \\ \wedge 2 \end{array} = \begin{array}{l} V \text{ NØ} \\ \$ \text{ RI(01,02,39,} \\ \quad 40,41,42) \end{array}$$

$$\begin{array}{l} V \text{ NØ} \\ + \text{ NU(P)} \\ \wedge 2 \end{array} = \begin{array}{l} V \text{ NØ} \\ \$ \text{ RI(43,04,03,} \\ \quad 07,34,44,45,46) \end{array}$$

APPENDIX B

42a

- A = ADJECTIVE OR ADVERB (PHRASE) [VALUE OF CONJUNCTION FEATURE #CONNECTS OR INTRODUCES# (C-I)]
- = ADJECTIVE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- = ADVERSATIVITY [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS, CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- = ANTE (= PRE-POSED) RELEVANT FOR MODIFIERS OF NP, A, AV, OR NU ONLY [VALUE FOR ADVERB FEATURE #POSITION# (POS)]
- AG = ACCUSATIVE [VALUE FOR ADJECTIVE FEATURE #FORM OF OBJECT# (OR)]
- = ACCUSATIVE [VALUE FOR ADVERB FEATURES #REQUIRES COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
- = ACCUSATIVE [VALUE FOR VERB FEATURE #DEEP SUBJECT# (DS)]
- = ACCUSATIVE [VALUE FOR VERB FEATURE #OBJECT OR COMPLEMENT SYNTAX# (OB)]
- AB = ABSTRACT [VALUE FOR NOUN FEATURE #TYPE# (TY), FOR VERB FEATURE #TYPE OF SUBJECT# (TS), FOR VERB AND ADJECTIVE FEATURE #TYPE OF OBJECT# (TO)]
- = ABSTRACT [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- AC = ADJECTIVE COMPLEMENT (FOR SENSORY VERBS, E.G., 'SMELL GOOD') [VALUE FOR VERB OR ADJECTIVE FEATURE #REQUIRED ADVERBIALS# (RA)]
- = ADJECTIVE COMPLEMENT [VALUE FOR ADVERB FEATURES #REQUIRES COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
- = TAKES ADJ COMPLEMENT WITHOUT 'BE' [VALUE FOR VERB FEATURE #OBJECT OR COMPLEMENT SYNTAX# (OB)]
- AL = ANIMAL [VALUE FOR NOUN FEATURE #TYPE# (TY), FOR VERB FEATURE #TYPE OF SUBJECT# (TS), FOR VERB AND ADJECTIVE FEATURE #TYPE OF OBJECT# (TO)]
- = ANIMAL [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- ALS = 'WHEN' (E.G., 'THE TIME WHEN I LIVED THERE') [VALUE FOR NOUN FEATURE #RELATIVE ADVERB# (RL)]
- AN = ANIMATE [VALUE FOR NOUN FEATURE #TYPE# (TY), FOR VERB FEATURE #TYPE OF SUBJECT# (TS), FOR VERB AND ADJECTIVE FEATURE #TYPE OF OBJECT# (TO)]

AN = ANIMATE [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF
 COMPLEMENT# (TC)]
 APP = APPROXIMATION [SPECIFICATION OF ADVERB-CONJUNCTION-
 PREPOSITION VALUE #DEGREE# (CD)]
 AV = ADVERB (INCLUDING PRPH) [VALUE FOR ADVERB FEATURE
 #MODIFIERS# (MD)]
 = ADVERB [VALUE FOR PREPOSITION FEATURE #REQUIRES COMPLEMENT#
 (RC)]
 = ANY TYPE OF ADVERB [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV DIR = ADVERB OF DIRECTION TO [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV DUR = ADVERB OF DURATION [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV LOC = ADVERB OF LOCATION [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV MAN = ADVERB OF MANNER [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV MSR = ADVERB OF MEASURE [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV ORN = ADVERB OF ORIGIN (DIRECTION FROM) [VALUE FOR ADVERB FEATURES
 #REQUIRES COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV PLC = ADVERB OF PLACE [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV PNC = ADVERB OF PUNCTUALITY [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 AV TIM = ADVERB OF TIME [VALUE FOR ADVERB FEATURES #REQUIRES
 COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
 BC = TAKES 'BE' + NP OR ADJ [VALUE FOR VERB FEATURE/OBJECT OR
 COMPLEMENT SYNTAX# (OB)]
 BP = BODY PART [VALUE FOR NOUN FEATURE#TYPE# (TY), FOR VERB
 FEATURE#TYPE OF SUBJECT# (TS), FOR VERB AND ADJECTIVE
 FEATURE#TYPE OF OBJECT# (TO)]
 = BODY PART [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF
 COMPLEMENT# (TC)]
 C = COMPARATIVE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
 = 'SO+ADJ+THAT' OR 'TO A DEGREE THAT' PARAPHRASE POSSIBLE

(CONSECUTIVE) [VALUE FOR ADVERB FEATURE #PARENTHETICAL# (PA)]

- CA = CAUSE [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS, CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- CC = CONCESSIVE [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS, CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- CD = CONDITION [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS, CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- CL = MAIN CLAUSE (SUBJUNCTIVE IN GERMAN) [VALUE FOR VERB FEATURE #OBJECT OR COMPLEMENT SYNTAX#(OB)]
- CLG = MAIN CLAUSE, AS IN 'DIE BEHAUPTUNG, DIES SEI DIE WAHRHEIT' [VALUE FOR NOUN FEATURE#ATTRIBUTIVE#(TA)]
- CM = TAKES OPTIONAL 'BE' + NP OR ADJ (E.G., 'THINK') [VALUE FOR VERB FEATURE#OBJECT OR COMPLEMENT SYNTAX#(OB)]
- CN = COUNT (ABSTRACT COUNTABLE NOUNS) [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
 = COUNT (ABSTRACT COUNTABLE NOUNS, E.G., 'IDEA') [VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB FEATURE#TYPE OF SUBJECT#(TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT#(TO)]
- CO = COLLECTIVE (COMPONENTS ARE COUNTABLE) [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
 = COLLECTIVE (COMPONENTS MAY BE COUNTED; CAN BE USED WITH THE VERB 'DISPERSE'; E.G., 'GROUP, HERD, GOVERNMENT') [VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB FEATURE#TYPE OF SUBJECT#(TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT#(TO)]
- COM = COMPARISON [SPECIFICATION OF ADVERB-CONJUNCTION-
 PREPOSITION VALUE #MODAL# (M)]
 = COMPARISON [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION
 VALUE #DEGREE# (CD)]
- COM PEJ= COMPARISON PEJORATIVE [SPECIFICATION OF ADVERB-CONJUNCTION-
 PREPOSITION VALUE #MODAL# (M)]
- CONJ = CONJUNCTIVE (TAKES PL. VERB: 'AND') [VALUE OF CONJUNCTION
 FEATURE #TYPE# (TY)]
- C-I = CONNECTS OR INTRODUCES (CLAUSES, NOUN PHRASES, VERB PHRASES,
 ETC.) [CONJUNCTION FEATURE]

- D = DECLARATIVE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- = DEGREE [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS (CONJUNCTIONS, AND PREPOSITIONS); VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- D_G = DATIVE [VALUE FOR ADJECTIVE FEATURE#FORM OF OBJECT#(OB)]
- = DATIVE [VALUE FOR ADVERB FEATURES #REQUIRES COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
- = DATIVE [VALUE FOR VERB FEATURE#DEEP SUBJECT#(DS)]
- = DATIVE [VALUE FOR VERB FEATURE#OBJECT OR COMPLEMENT SYNTAX# (OB)]
- DEF = DEFINITE [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS (CONJUNCTIONS, AND PREPOSITIONS); VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- DF = DERIVED FROM [NOUN FEATURE]
- DIR = DIRECTION TO [VALUE FOR VERB OR ADJECTIVE FEATURE#REQUIRED ADVERBIALS#(RA)]
- = DIRECTIONAL ADVERBIAL COMPLEMENT (E.G., 'A TRIP ACROSS EUROPE') [VALUE FOR NOUN FEATURE#ATTRIBUTIVE#(TA)]
- DISJ = DISJUNCTIVE (TAKES SG. VERB: 'OR') [VALUE OF CONJUNCTION FEATURE #TYPE# (TY)]
- DI-F = DIRECTION FROM [SPECIFICATION OF ADVERB-CONJUNCTION- PREPOSITION VALUE #LOCATION# (L)]
- DI-T = DIRECTION TO [SPECIFICATION OF ADVERB-CONJUNCTION- PREPOSITION VALUE #LOCATION# (L)]
- DOR = DIRECTION OR ORIGIN (ADVERB OF DIRECTIONALITY) [VALUE FOR VERB FEATURE#OPTIONAL ADVERBIALS#(OA)]
- DS_G = DEEP SUBJECT [VERB FEATURE INDICATED ONLY IF THE DEEP SUBJECT DOES NOT OCCUR AS A NOMINATIVE IN THE SURFACE SENTENCE]
- DU = DURATION (TIME SPAN ANSWERING THE QUESTION 'HOW LONG?', E.G., 'FOR ELEVEN DAYS') [SPECIFICATION OF ADVERB- CONJUNCTION- PREPOSITION VALUE #TIME# (T)]
- DUR = DURATIONAL [VALUE FOR VERB OR ADJECTIVE FEATURE#REQUIRED ADVERBIALS#(RA)]
- E = ANY TYPE OF NOUN [VALUE FOR ADJECTIVE FEATURE#TYPE OF OBJECT# (TO)]
- = 'ENTIA' (ANY TYPE OF NOUN) [VALUE FOR VERB FEATURE#TYPE OF OBJECT#(TO)]

- E = 'ENTIA' (ANY TYPE OF NOUN) [VALUE FOR VERB FEATURE#TYPE OF SUBJECT#(TS)]
- = EQUATIVE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- EV = EVALUATION OF SUBJECT [SPECIFICATION OF ADVERB-CONJUNCTION-
PREPOSITION VALUE #MODAL# (M)]
- EXT = EXTENDED (TIME SPAN ANSWERING THE QUESTION 'WHEN?', E.G.,
'TODAY') [SPECIFICATION OF ADVERB-CONJUNCTION-
PREPOSITION VALUE #TIME# (T)]
- F = SENTENCE-FINAL (RELEVANT FOR MODIFIERS OF S AND V ONLY)
[VALUE FOR ADVERB FEATURE #POSITION# (POS)]
- FE = FEMALE [VALUE FOR THE NOUN FEATURE#SEX#(SX)]
- FM = FORM OF ADJECTIVE [ADJECTIVE FEATURE]
- = FORM (FOR NOMINALIZED ADJECTIVES) [NOUN FEATURE]
- FR = FREQUENCY (REPETITIVE) [SPECIFICATION OF ADVERB-CONJUNCTION
-PREPOSITION VALUE #TIME# (T)]
- FS = SYNTACTIC FORM OF SUBJECT [REQUIRED FEATURE OF VERB UNLESS
IT ALLOWS ONLY A NOUN PHRASE AS SUBJECT]
- FT = 'FOR-TO' COMPLEMENT [VALUE FOR ADVERB FEATURES #REQUIRES
COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
- FT_E = 'FOR-TO' COMPLEMENT [VALUE FOR VERB FEATURE#FORM OF SUBJECT#
(FS)]
- FU = FUTURE [SPECIFICATION OF ADVERB-CONJUNCTION-
PREPOSITION VALUE #TIME# (T)]
- = OCCURS WITH VERBS IN FUTURE [VALUE FOR ADVERB FEATURE
#TENSE# (TN)]
- GG = GENITIVE [VALUE FOR ADJECTIVE FEATURE#FORM OF OBJECT#(OR)]
- = GENITIVE [VALUE FOR ADVERB FEATURES #REQUIRES COMPLEMENT#
(RC) AND #OPTIONAL COMPLEMENT# (OC)]
- = GENITIVE [VALUE FOR VERB FEATURE#DEEP SUBJECT#(DS)]
- = GENITIVE [VALUE FOR VERB FEATURE#OBJECT OR COMPLEMENT SYNTAX#
(OB)]
- GR = GERUND [VALUE FOR ADVERB FEATURES #REQUIRES COMPLEMENT# (RC)
AND #OPTIONAL COMPLEMENT# (OC)]
- GR_E = GERUND [VALUE FOR VERB FEATURE#FORM OF SUBJECT#(FS)]
- H = 'IT - HOW' PARAPHRASE POSSIBLE [VALUE FOR ADVERB FEATURE

#PARENTHETICAL# (PA)]

- HU = HUMAN [VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB FEATURE #TYPE OF SUBJECT#(TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT#(TO)]
- = HUMAN [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- H-S = HIGHER SCALE [SPECIFICATION OF ADVERB-CONJUNCTION-
PREPOSITION VALUE #DEGREE# (CD)]
- I = IMPERATIVE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- = INSTRUMENTAL [VALUE FOR FEATURE #TYPE# (TY) USED WITH
ADVERBS, CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR
PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- = 'IT - THAT' PARAPHRASE POSSIBLE [VALUE FOR ADVERB FEATURE
#PARENTHETICAL# (PA)]
- = SENTENCE-INITIAL (RELEVANT FOR MODIFIERS OF S AND V ONLY)
[VALUE FOR ADVERB FEATURE #POSITION# (POS)]
- = UNMARKED INFINITIVE [VALUE FOR VERB FEATURE#OBJECT OR
COMPLEMENT SYNTAX#(OB)]
- ICL = INTERROGATIVE CLAUSE [VALUE FOR ADVERB FEATURES #REQUIRES
COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
- = INTERROGATIVE CLAUSE [VALUE FOR VERB FEATURE#FORM OF SUBJECT#
(FS)]
- II_G = INTERROGATIVE ADVERB + UNMARKED INFINITIVE [VALUE FOR VERB
FEATURE#FORM OF SUBJECT#(FS)]
- IMI = INTERROGATIVE ADVERB + MARKED INFINITIVE [VALUE FOR ADVERB
FEATURES #REQUIRES COMPLEMENT# (RC) AND #OPTIONAL
COMPLEMENT# (OC)]
- IMI_E = INTERROGATIVE ADVERB + MARKED INFINITIVE [VALUE FOR VERB
FEATURE#FORM OF SUBJECT#(FS)]
- IN = INANIMATE [VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB
FEATURE#TYPE OF SUBJECT#(TS), FOR VERB AND ADJECTIVE
FEATURE#TYPE OF OBJECT#(TO)]
- = INANIMATE [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF
COMPLEMENT# (TC)]
- INC = INCIPIENT [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION
VALUE #TIME# (T)]
- IND = INDEFINITE [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS

(CONJUNCTIONS, AND PREPOSITIONS); VALUE FOR PREPOSITION
FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]

INST = INSTANTANEOUS (POINT IN TIME, E.G., 'AT 8 P.M.')
[SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE
#TIME# (T)]

IT = 'IT' [VALUE FOR VERB FEATURE#FORM OF SUBJECT#(FS)]

L = LOCATION [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS
(CONJUNCTIONS, AND PREPOSITIONS); VALUE FOR PREPOSITION
FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]

LS = LOWER SCALE [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION
VALUE #DEGREE# (CD)]

M = MODAL [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS
(CONJUNCTIONS, AND PREPOSITIONS); VALUE FOR PREPOSITION
FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]

= SENTENCE-MEDIAL (RELEVANT FOR MODIFIERS OF S AND V ONLY)
[VALUE FOR ADVERB FEATURE #POSITION# (POS)]

MA = MACHINE (SINCE THEY CAN PERFORM SOME HUMAN ACTIVITIES)
[VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB FEATURE#TYPE OF
SUBJECT#(TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT#
(TD)]

= MACHINE [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF
COMPLEMENT# (TC)]

= MALE [VALUE FOR THE NOUN FEATURE SEX (SX)]

MAN = MANNER [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION
VALUE #MODAL# (M)]

= MANNER [VALUE FOR VERB OR ADJECTIVE FEATURE#REQUIRED
ADVERBIALS#(RA)]

MC = MAIN CLAUSE [VALUE OF CONJUNCTION FEATURE #CONNECTS OR
INTRODUCES# (C-I)]

MD = MODIFIES NOUNS OF THE SPECIFIED TYPE [MANDATORY ADJECTIVE
FEATURE]

= MODIFIES (THE ADVERB MAY MODIFY VERBS, SENTENCES, OR NP'S)
[MANDATORY ADVERB FEATURE]

ME = MEASURE [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS,
CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR PREPOSITION
FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]

MED = MEDIUM [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION
VALUE #DEGREE# (CD)]

MI = MARKED INFINITIVE [VALUE FOR ADVERB FEATURES #REQUIRES

COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]

- MI = MARKED INFINITIVE [VALUE FOR VERB FEATURE#FORM OF SUBJECT# (FS)]
- MO = MODALITY [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS, CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- MOD = MODE OF EXISTENCE [SPECIFICATION OF ADVERB-CONJUNCTION- PREPOSITION VALUE #MODAL# (M)]
- MS = MASS (HOMOGENEOUS) [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- = MASS (HOMOGENEOUS; MAY OCCUR WITHOUT ARTICLE IN THE SINGULAR ; E.G. 'MILK, SAND') [VALUE FOR NOUN FEATURE#TYPE# (TY), FOR VERB FEATURE#TYPE OF SUBJECT# (TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT# (TO)]
- MSR = MEASURABLE (E.G., 'WIDE' OR 'STRONG' AS IN 'FIVE INCHES WIDE', 'SEVEN MEN STRONG') [VALUE FOR ADJECTIVE FEATURE#TYPE# (TY)]
- = MEASURE [VALUE FOR VERB OR ADJECTIVE FEATURE#REQUIRED ADVERBIALS# (RA)]
- N = NEGATED D, Q, I, OR S [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- = NOUN (PHRASE) [VALUE OF CONJUNCTION FEATURE #CONNECTS OR INTRODUCES# (C-I)]
- NA = TAKES NP OR ADJ COMPLEMENT WITHOUT 'BE' [VALUE FOR VERB FEATURE#OBJECT OR COMPLEMENT SYNTAX# (OB)]
- NC = TAKES NP COMPLEMENT WITHOUT 'BE' (E.G., 'ELECT') [VALUE FOR VERB FEATURE#OBJECT OR COMPLEMENT SYNTAX# (OB)]
- NG = THE VERB DOES NOT FORM THE PROGRESSIVE [VALUE FOR REQUIRED VERB FEATURE#TYPE# (TY)]
- NM = PROPER NAME [VALUE FOR NOUN FEATURE#TYPE# (TY), FOR VERB FEATURE#TYPE OF SUBJECT# (TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT# (TO)]
- = PROPER NAME [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- NP = NOUN PHRASE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- = NOUN PHRASE [VALUE FOR PREPOSITION FEATURE #REQUIRES COMPLEMENT# (RC)]
- = NOUN PHRASE [VALUE FOR VERB FEATURE#FORM OF SUBJECT# (FS)]

- NP = THE VERB DOES NOT PASSIVIZE; VERBS MARKED VI OR VR DO NOT NEED THIS DESCRIPTOR. [VALUE FOR REQUIRED VERB FEATURE #TYPE# (TY)]
- NP_G = NOUN PHRASE [VALUE FOR ADVERB FEATURES #REQUIRES COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]
- NU = NUMBERS [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- O_E = NOUN PHRASE (NP) AS OBJECT [VALUE FOR VERB FEATURE #OBJECT OR COMPLEMENT SYNTAX# (OB)]
- OA = OPTIONAL ADVERBIALS
- OB = FORM OF OBJECT [ADJECTIVE FEATURE]
= OBJECT (IN CASE OF DEVERBATIVE NOUNS, AS E.G., 'DEPENDENCE ON') [NOUN FEATURE]
= SYNTACTIC FORM OF OBJECT(S) OR COMPLEMENT(S) [VERB FEATURE]
= 'WHETHER' (E.G., 'THE QUESTION WHETHER THIS IS SO') [VALUE FOR NOUN FEATURE #RELATIVE ADVERB# (RL)]
- OC = OPTIONAL COMPLEMENT [ADVERB FEATURE]
- ORN = ORIGIN (DIRECTION FROM) [VALUE FOR VERB OR ADJECTIVE FEATURE #REQUIRED ADVERBIALS# (RA)]
- P = PARENTHETICAL [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS (CONJUNCTIONS, AND PREPOSITIONS); VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
= PLURAL NOUN ONLY [VALUE FOR VERB FEATURE TYPE OF OBJECT (TO)]
= PLURAL NOUN ONLY [VALUE FOR VERB FEATURE #TYPE OF SUBJECT# (TS)]
= PLURAL (I.E., THE ADVERB REQUIRES A PLURAL SUBJECT OR A SINGULAR SUBJECT WITH A 'MIT-PRPH') [VALUE FOR ADVERB FEATURE #TYPE OF SUBJECT# (TS)]
= POSITIVE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
= POST (= POST-POSED) RELEVANT FOR MODIFIERS OF NP, A, AV, OR NU ONLY [VALUE FOR ADVERB FEATURE #POSITION# (POS)]
- PA = OCCURS WITH VERBS IN PAST TENSE [VALUE FOR ADVERB FEATURE #TENSE# (TN)]
= PARAPHRASABILITY (RELEVANT ONLY FOR PARENTHETICAL ADVERBS) [ADVERB FEATURE]
= PAST [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE

*TIME# (T)]

PAPL = PAST PARTICIPLE [VALUE FOR ADJECTIVE FEATURE#FORM#(FM)]
 = PAST PARTICIPLE [VALUE FOR VERB FEATURE#OBJECT OR COMPLEMENT SYNTAX#(OB)]

PF = OCCURS WITH VERBS IN ANY PERFECT TENSE [VALUE FOR ADVERB FEATURE #TENSE# (TN)]

PL = PLANT [VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB FEATURE #TYPE OF SUBJECT#(TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT#(TO)]
 = PLANT [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]

PLC = PLACE (LOCATIVE OR DIRECTIONAL) [VALUE FOR VERB OR ADJECTIVE FEATURE#REQUIRED ADVERBIALS#(RA)]

PLU = PLURAL, MASS, OR COLLECTIVE NOUN [VALUE FOR ADJECTIVE FEATURE#TYPE OF NOUN MODIFIED#(MD)]

PNC = PUNCTUAL [VALUE FOR VERB OR ADJECTIVE FEATURE#REQUIRED ADVERBIALS#(RA)]

PO = PHYSICAL OBJECT [VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB FEATURE#TYPE OF SUBJECT#(TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT#(TO)]
 = PHYSICAL OBJECT [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]

POS = POSITION (PRE- OR POST-POSED; SENTENCE INITIAL, MEDIAL, OR FINAL) [ADVERB FEATURE]
 = POSITION (PRE- OR POST-POSED) [PREPOSITION FEATURE]

POST = POST-POSED TO THE NP OR AV [VALUE FOR PREPOSITION FEATURE #POSITION# (POS)]

PO-T = POSTERIOR TO [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #TIME# (T)]

PP = PURPOSE [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS, CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]

PR = OCCURS WITH VERBS IN PRESENT TENSE [VALUE FOR ADVERB FEATURE #TENSE# (TN)]
 = PRESENT [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #TIME# (T)]

- PRE = PRE-POSED TO THE NP OR AV [VALUE FOR PREPOSITION FEATURE #POSITION# (POS)]
- PRPL = PRESENT PARTICIPLE [VALUE FOR ADJECTIVE FEATURE#FORM#(FM)]
- PR-T = PRIOR TO [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #TIME# (T)]
- PU = PUNCTUAL [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #TIME# (T)]
- Q = QUESTION [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- QU = QUANTITY (-- + (OF) NP; E.G., 'GROUP, GLASS, HALF, AS IN A 'GLASS OF MILK') [VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB FEATURE#TYPE OF SUBJECT#(TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT#(TO)]
- = QUANTITY (--+ (OF) NP; E.G., 'GROUP, AS IN A 'GROUP OF ITEMS') [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- R = REFLEXIVE [VALUE FOR VERB FEATURE#TYPE OF OBJECT#(TO)]
- = RESULT [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS, CONJUNCTIONS, AND PREPOSITIONS; VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- RA = REQUIRED ADVERBIALS [VERB FEATURE]
- = REQUIRES AN ADVERB (E.G., 'RESIDENT' AS PREDICATIVE ADJ.) [ADJECTIVE FEATURE]
- RC = REQUIRES COMPLEMENT (ADVERBS, CLAUSES OR PHRASES) [ADVERB FEATURE]
- = REQUIRES COMPLEMENT (NOUN PHRASE OR ADVERB) [PREPOSITION FEATURE]
- RCC = RECIPROCAL [VALUE FOR VERB FEATURE#TYPE OF OBJECT#(TO)]
- RES = RESTRICTIVE [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #MODAL# (M)]
- RL = RELATIVE ADVERB (FOR DEVERBATIVE NOUNS) [NOUN FEATURE]
- S = SENTENCE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- = SOCIATIVE [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS, CONJUNCTIONS, AND PREPOSITIONS, VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- SC = SUBORDINATE CLAUSE [VALUE OF CONJUNCTION FEATURE #CONNECTS OR INTRODUCES# (C-I)]

SE = SEQUENTIAL (SE WITHOUT INC OR TRM MEANS -SEQUENTIAL BUT NOT INITIAL OR FINAL-, E.G., 'SECONDLY') [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #TIME# (T)]

SIM = SIMULTANEOUS WITH [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #TIME# (T)]

SM = STATE OF MIND [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #MODAL# (M)]

SP = SUPERLATIVE [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]

STA = STATIC [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #LOCATION# (L)]

SX = SEX [NOUN FEATURE]

T = TIME [VALUE FOR FEATURE #TYPE# (TY) USED WITH ADVERBS (CONJUNCTIONS, AND PREPOSITIONS); VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]

TA = TAKES ATTRIBUTE [NOUN FEATURE]

TC = SEMANTIC TYPE OF COMPLEMENT [PREPOSITION FEATURE]

TH = 'THAT'-CLAUSE (NON-RELATIVE 'THAT'-CLAUSES; E.G., 'HIS CLAIM THAT THIS WAS SO') [VALUE FOR NOUN FEATURE#ATTRIBUTIVE#(TA)]

= 'THAT'-CLAUSE [VALUE FOR ADJECTIVE FEATURE#TYPE OF NOUN MODIFIED#(MD)]

= 'THAT'-CLAUSE [VALUE FOR ADVERB FEATURES #REQUIRES COMPLEMENT# (RC) AND #OPTIONAL COMPLEMENT# (OC)]

= 'THAT'-CLAUSE [VALUE FOR VERB FEATURE#FORM OF SUBJECT#(FS)]

TIM = TIME (PUNCTUAL OR DURATIONAL) [VALUE FOR VERB OR ADJECTIVE FEATURE#REQUIRED ADVERBIALS#(RA)]

TM = THE ADJECTIVE MAY UNDERGO "TOUGH MOVEMENT" (E.G., 'HARD, EASY') [VALUE FOR ADJECTIVE FEATURE#TYPE#(TY)]

TN = TENSE (THE ADVERB REQUIRES THAT THE VERB OCCUR IN A SPECIFIC TENSE(S); THIS SUBSCRIPT IS NOT CODED IF THE SAME INFORMATION IS CONTAINED UNDER TY IN ONE OF THE VALUES PR, PA OR FU) [ADVERB FEATURE]

TO = SEMANTIC TYPE OF OBJECT [ADJECTIVE FEATURE]

= SEMANTIC TYPE OF OBJECT [NOUN FEATURE]

= SEMANTIC TYPE OF OBJECT [VERB FEATURE]

TRM = TERMINATING [SPECIFICATION OF ADVERB-CONJUNCTION-PREPOSITION VALUE #TIME# (T)]

- TS = SEMANTIC TYPE OF SENTENCE SUBJECT REQUIRED (RELEVANT ONLY WITH ADVERBS MODIFYING VERBS AND, POSSIBLY, SENTENCES) [ADVERB FEATURE]
- TS = SEMANTIC TYPE OF SUBJECT [REQUIRED FEATURE OF VERB]
- TV = SEMANTIC TYPE OF VERB WITH WHICH THE ADVERB MAY BE USED (RELEVANT ONLY WITH ADVERBS MODIFYING VERBS) [ADVERB FEATURE]
- TY = SEMANTIC TYPE OF ADVERB [MANDATORY ADVERB FEATURE]
- = SEMANTIC TYPE OF PREPOSITION [PREPOSITION FEATURE]
- = TYPE OF ADJECTIVE [ADJECTIVE FEATURE]
- = TYPE OF CONJUNCTION [CONJUNCTION FEATURE]
- = TYPE OF NOUN [MANDATORY NOUN FEATURE]
- = TYPE OF VERB (TRANSITIVITY) [REQUIRED FEATURE OF VERB]
- UN = UNIT (ADV = QUANT + --; E.G., 'MILE, YEAR', AS IN 'FIVE MILES LONG, TO WAIT TWO YEARS') [VALUE FOR NOUN FEATURE#TYPE#(TY), FOR VERB FEATURE#TYPE OF SUBJECT#(TS), FOR VERB AND ADJECTIVE FEATURE#TYPE OF OBJECT#(TO)]
- = UNIT (ADV = QUANT + --; E.G., 'MILE', AS IN 'FIVE MILES LONG') [VALUE FOR PREPOSITION FEATURE #SEMANTIC TYPE OF COMPLEMENT# (TC)]
- V = VERB (PHRASE) [VALUE OF CONJUNCTION FEATURE #CONNECTS OR INTRODUCES# (C-I)]
- = VERB [VALUE FOR ADVERB FEATURE #MODIFIERS# (MD)]
- VI = INTRANSITIVE [VALUE FOR REQUIRED VERB FEATURE#TYPE#(TY)]
- VR = TAKES AN OBJECT WHICH MUST BE REFLEXIVE [VALUE FOR REQUIRED VERB FEATURE#TYPE#(TY)]
- VT = TAKES AT LEAST ONE OBJECT WHICH IS NOT A REFLEXIVE PRONOUN [VALUE FOR REQUIRED VERB FEATURE#TYPE#(TY)]
- VTC = TAKES A COGNATE OBJECT ONLY, E.G., 'TO DANCE A WALTZ' [VALUE FOR REQUIRED VERB FEATURE#TYPE#(TY)]
- VT,VR = TAKES AT LEAST 2 OBJECTS, ONE OF WHICH MUST BE REFLEXIVE AND ONE WHICH IS NOT [VALUE FOR REQUIRED VERB FEATURE#TYPE#(TY)]
- W = POST-SENTENTIAL WHICH-RELATIVE CLAUSE POSSIBLE [VALUE FOR ADVERB FEATURE #PARENTHETICAL# (PA)]
- WARUM = 'WHY' (E.G., 'THE REASON WHY HE DID IT') [VALUE FOR NOUN FEATURE

RELATIVE ADVERB (RL)]

- WIE = 'HOW' (E.G., 'THE QUESTION OF HOW THIS HAPPENED') [VALUE FOR NOUN FEATURE#RELATIVE ADVERB#(RL)]
- WO = 'WHERE' (E.G., 'THE PLACE WHERE I SAW YOU') [VALUE FOR NOUN FEATURE#RELATIVE ADVERB#(RL)]
- WOHIN = 'WHERE TO' (E.G., 'THE TOWN WHERE YOU WENT') [VALUE FOR NOUN FEATURE#RELATIVE ADVERB#(RL)]
- ZU = MARKED INFINITIVE (E.G., 'ATTEMPT', AS IN 'THE ATTEMPT TO DO SOMETHING') [VALUE FOR NOUN FEATURE#ATTRIBUTIVE#(TA)]
- * = NOT
- ? = SPECIAL SYMBOL MARKING COMPONENT PROBABLY REQUIRING ADDITIONAL ATTENTION IN THE FUTURE

REFERENCES

1. Hornby, A.S., E.V. Gatenby, and H. Wakefield. 1963. *The Advanced Learner's Dictionary of Current English*. 2nd ed. London: Oxford University Press.
2. Schmidt, Veronika. 1968. *Die Streckformen des deutschen Verbums: Substantivisch-verbale Wortverbindungen in publizistischen Texten der Jahre 1948 bis 1967*. Halle (Saale): VEB Max Niemeyer Verlag.
3. *Webster's Seventh New Collegiate Dictionary*. 1967. Springfield, Mass.: G. & C. Merriam.
4. Wildhagen, Karl, and Will Heraucourt. 1953, 1959. *English-German German-English Dictionary*. 2 vols. London: George Allen & Unwin; Wiesbaden: Brandstetter.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) University of Texas at Austin Linguistics Research Center Austin, TX 78712		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP N/A	
3. REPORT TITLE DEVELOPMENT OF GERMAN-ENGLISH MACHINE TRANSLATION SYSTEM			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Annual Report 1 February 1971 - 31 January 1972			
5. AUTHOR(S) (First name, middle initial, last name) Dr. Winfred P. Lehmann Dr. Rolf A. Stachowitz			
6. REPORT DATE April 1972	7a. TOTAL NO. OF PAGES 97	7b. NO. OF REFS 4	
8a. CONTRACT OR GRANT NO. F30602-70-C-0118 Job Order No. 45940000 68 003 62702F	9a. ORIGINATOR'S REPORT NUMBER(S) None		
	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) RADC-TR-72-74		
10. DISTRIBUTION STATEMENT Approved for public release; distribution unlimited.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY Rome Air Development Center (IRDT) Griffiss Air Force Base, New York 13440	
13. ABSTRACT This report deals with advances in linguistic analysis and programming. The theoretical theses for this work were presented in the first annual report. This second report accordingly concentrates on progress in the descriptive analysis of German and English, and on computer programs which have been developed during the year. Linguistic work during the period was directed at expanding the dictionaries for both German and English. The number of lexical items coded was increased, as were the features associated with each item. The coding involved the linguists in some of the most complex problems of linguistic description, such as treatment of adverbs and extended forms for verbs. Since adverbs had not been adequately classified in any existing grammars of German or English, a new classificatory schema is being developed. The extended forms of German verbs have been treated in a recent monograph, which has served as the basis for the lexical analysis of these constructions within the German-English MT System. These extended forms are particularly difficult in having a small group of verbs used primarily to depict the verbal component and nouns combined with them to carry the semantic portion of the expression. Imaginative programs have enabled us to identify the specific problems encountered in our texts. Others among those reported here are representative of the massive set of programs necessary to manage the deep as well as the surface structures of language.			

DD FORM 1 NOV 65 1473

102

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14.	KEY WORDS	LINK A		LINK B		LINK C	
		ROLE	WT	ROLE	WT	ROLE	WT
	Documentation, Machine Translation Social Science, Language						

SAC--Griffiss AFB NY

UNCLASSIFIED

Security Classification